December 1996

Selection and Prediction

Using Latent Variable Models

1995

Li-Chiao Huang

Bengt Muthén

Graduate School of Education & Information Studies

UCLA

Los Angeles, CA 90024-1521

This research was supported by grant no. N0014-93-1-0619 from the U.S. Navy/Office of Naval Research.

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

Contents

Abstract vi

CHAPTERS

1 Introduction 1

2 Literature Review 8
2.1 Factor Score Determinacy and Factor Score Estimation 8
2.2 Selection and Classification 11
2.3 Differential Assignment Theory 14
2.4 Predictive Validity and Cross-Validity 16
2.4.1 Predictive Validity in Relation to *g* and *s* 16
2.4.2 Cross-Validity 19

3 Methodology 22
3.1 Technical Definitions 22
3.1.1 Measurement Model 22
3.1.2 Predictive Validity 23
3.1.3 Regression Method of Factor Score Estimation 24
3.1.4 Factor Score Determinacy 24
3.1.5 Sensitivity, Specificity, and Proportion of Correct Decisions 25
3.1.6 The Spearman Rank-Order Correlation Coefficient 26
3.1.7 Cross-Validation 26
3.2 The Monte Carlo Study 27
3.2.1 Design of the Monte Carlo Study 29
3.2.2 Analyses 34

4 Results of the Monte Carlo Study 36
4.1 Results of Sensitivity, Specificity, and Proportion of Correct Decisions 36
4.2 Results of Spearman Rank-Order Correlations 39
4.2.1 Applicant Sample 39
4.2.2 Cross-Validation Sample 40
4.3 Effects of Selection Methods 41
4.4 Effects of Selection Ratio and Sample Size 41
4.5 Effects of Models 43
4.5.1 R Square 43
4.5.2 Standardized Regression Slopes 43
4.5.3 Factor Loadings 45

4.6 The Increment of Prediction From the Specific Factors 45
4.6.1 The Gains in Classification 45
4.6.2 The Gains in Spearman Rank-Order Correlation Coefficient 46

5 Real Data Application 48
5.1 The Army Project A and the Marine Corps JPM Data 48
5.1.1 Sample 49
5.1.2 Variables 49
5.2 Structural Models 50
5.3 Procedures of Analyses 52
5.4 Results 54
5.4.1 Factor Determinacy 54
5.4.2 Standardized Regression Equation and the Adjusted R Square 55
5.4.3 The Spearman Rank-Order Correlation Coefficient 56
5.4.4 Sensitivity, Specificity, and Proportion of Correct Decisions 57
5.4.5 Profile Description 59

6 Discussion 61
6.1 Discussion of the Monte Carlo Study 61
6.2 Discussion of the Real Data Application 62
6.3 Limitations, Implications, and Recommendations 65

Tables 70

Figures 93

Appendix 118
A The Factor Structure 118
B Job Profile Description 121

References 133

List of Tables

1 Design of the Simulation Study 70

2 Model Parameter Values 71

3 Summary of Classification and Spearman Rank Correlation for $g + 1\ s$ 74

4 Summary of Classification and Spearman Rank Correlation for $g + 2\ s$ 76

5 Summary of Classification and Spearman Rank Correlation for $g + 3\ s$ 78

6 Summary of Cross-Validation for $g + 1\ s$ 80

7 Summary of Cross-Validation for $g + 2\ s$ 81

8 Summary of Cross-Validation for $g + 3\ s$ 82

9 The Number of Incumbents in the Army Project A and the Marine Corps JPM Studies 83

10 Summary of Tests in the Studies 84

11 Factor Scores Description 85

12 Estimates for the Standardized Prediction Equation and Adjusted R Squares 89

13 Estimates for the Standardized Prediction Equation (Top 50%) 90

14 Spearman Rank-Order Correlations 91

15 Job Classifications 92

List of Figures

1 Observed Data Pattern with Missing Data on Criterion 93

2 General Form of the Decision Table 94

3 Data Pattern for Cross-Validation in the Selected Sample 95

4 Single-Factor Model 96

5 Hierarchical Factor Model 97

6 Data Pattern for the Artificial Data 98

7 Plots of Classification for $g + 1\,s$ 99

8 Plots of Classification for $g + 2\,s$ 102

9 Plots of Classification for $g + 3\,s$ 105

10 Plots of Difference for $g + 1\,s$ Classification 108

11 Plots of Difference for $g + 2\,s$ Classification 109

12 Plots of Difference for $g + 3\,s$ Classification 110

13 Plots of Spearman Rank-Order Correlation for $g + 1\,s$ 111

14 Plots of Spearman Rank-Order Correlation for $g + 2\,s$ 112

15 Plots of Spearman Rank-Order Correlation for $g + 3\,s$ 113

16 Plots of Difference for Cross-Validation of $g + 1\,s$ 114

17 Plots of Difference for Cross-Validation of $g + 2\,s$ 115

18 Plots of Difference for Cross-Validation of $g + 3\,s$ 116

19 Plot for 4% Increase in the Rate of Sensitivity 117

Abstract

The purpose of this study is to investigate the use of factor scores for improving predictive validity in personnel selection. Recent studies have shown that general ability is a good predictor of future performance. However, specific factors may be important indicators for personnel selection as well. In the latent variable context, the hierarchical model which considers general and specific factors simultaneously offers possibilities for solving prediction problems efficiently and parsimoniously. This type of modeling is useful in determining the relative importance of general and specific factors as predictors of the criterion variable, and for improving person-job match. Two approaches are used in the current investigation: an artificial data is used to study the prediction of future performance, and a real data application using the Army Project A and the Marine Corps JPM enlistment and performance data is used to study the practical implications of the use of specific factors.

The results of this study show that using specific factors in addition to the general factor as predictors provides better selection decisions. The illustration using the real data analyses suggests that including specific factors in predicting hands-on performance for most of the jobs under consideration creates gains in terms of sensitivity, specificity, and proportion of correct decisions.

CHAPTER 1

INTRODUCTION

Test scores are widely used as the basis for personnel decisions such as selection and placement of employees. For example, applicants for clerical positions must pass a typing test; colleges require a minimum score on the Scholastic Aptitude Test (SAT); government employees must take Civil Service examinations; and preschool children must meet minimum standards on intelligence tests for admission to many private kindergartens (Allred, 1991). In higher education, admission test scores are commonly used in decision-making regarding the admission of students. In specialized training programs, individuals may be admitted according to their performance on related aptitude tests. These testing programs provide an objective method of screening individuals to find those who are best qualified for the situation in question. Well-known selection tests in the United States include the Armed Services Vocational Aptitude Battery (ASVAB) for selection into military jobs and the General Aptitude Test Battery employment test for civilian jobs. In addition, the SAT, GRE, LSAT, and GMAT are used in the selection process for entry into colleges and universities.

Test scores should communicate some sense of how well a person can do a particular job or what aspects of the job a person can do well. Instead, scores on personnel selection tests currently reveal something about an examinee's relative standing with reference to all other examinees, which is useful for ranking applicants but is not very informative about how a person at any particular score level will perform a given

job. In order for such test scores to shed light on how a person will perform in a specific position, such measures of job competency would need to be referenced to some external scale of job requirements, not merely to the performance of other potential employees on the same test. The interpretation of a performance test score refers to the inferences about job performance that can be drawn from criterion test performance. Insofar as a criterion measure is representative of the work that is required on the job, some kind of inference is warranted from the test to the job domain. Thus, the process of investing a performance score with meaning begins with a careful study of the job and involves selecting tasks for testing that adequately represent the entire domain of the job's requirements.

To evaluate a selection test, we need to know whether or not the test is effective. The effectiveness of a testing program in selecting appropriate individuals depends on how well the test scores predict later performance on an objective performance measure. This is called predictive validity. Predictive validity indicates the extent to which an individual's future level on the criterion is predicted from prior test performance (Messick, 1989). A high degree of validity is necessary for obtaining a substantial benefit from using the test as a selection instrument (Cronbach & Gleser, 1965). The relationship between test scores and performance is often expressed as a validity coefficient that indicates how well the performance measure, or criterion, is predicted by the test. In other words, the measurement of performance should be evaluated before adopting any tests.

The use of factor scores as independent variables in multiple prediction has been advocated in the literature (e.g., Morris & Guertin, 1977; Hsu, 1991; Muthén & Hsu, 1993; Hsu, 1995). The superiority of factor scores as predictors was affirmed by Morris and Guertin's finding that factor scores led to a higher accuracy of prediction in a cross-validation sample than did the use of data-level variables. In addition, most latent-trait models have been used with tests in which performance is a function of one unobserved (latent) characteristic or trait, such as vocabulary level or mathematical reasoning ability (Lord & Novick, 1968). It is the goal of such models to estimate an examinee's standing on the continuous latent trait. In particular, factor score computation takes into account the relative contribution of each variable to the underlying trait as a function of the construct of a variable, as reflected in the size of the factor loading (Lord, 1980; Short, 1990; Muthén, Kao & Burstein, 1991). In other words, factors (or latent variables) may be viewed descriptively as representing the parsimonious rank reduction for observation obtained on a specific set of subjects on a given set of variables (Wackwitz & Horn, 1971). Therefore, factor scores should approximate the true score more closely than scores obtained by compound measures.

Two approaches, psychometric and econometric, have different priorities for evaluating applicants. In personnel work, the traditional, psychometric, approach has been used to rank applicants in order of their predicted performance. Since higher test scores indicate a probability of better performance on the job, the simple answer offered by psychological measurement is that higher is better. From this point of view, job performance is the essential consideration, and the costs of increased performance are not

calculated. On the other hand, because higher-aptitude personnel costs more to recruit, management specialists tend to use an econometric approach to questions of selection standards. In a military context, models for setting enlistment standards have recently been designed to locate the most cost-effective cutoff score (Zeidner & Johnson, 1991a, 1991b, 1991c). These models, however, do not help to answer the question of how much performance is sufficient. Instead, they set enlistment standards in order to minimize personnel costs per unit of productivity. From this econometric point of view, cost is the essential consideration (Wigdor & Green, 1991). Neither of these two approaches, the psychometric nor the econometric, is singularly satisfactory in evaluating applicants. If the military is to fulfill its mission to provide for the national defense, enlistment standards must be set at a level in which to achieve the recruiting goals (Hogan & Harris, 1994). Military and civilian policy-makers are interested in understanding the method by which qualification standards and recruit quality goals are established. Efficient analytical tools will provide a greater understanding of the distribution of recruit quality needed to maintain adequate levels of job performance. Therefore, the policy-making process would benefit from an approach which analyze both performance and cost (Green & Mavor, 1994).

General cognitive ability theory suggests that one general ability factor underlies all specific cognitive abilities. Proponents of this theory believe that the underlying variable, *g*, causes specific aptitudes to have validity in predicting job performance. If it is true that a single factor (*g*) underlies specific aptitudes, and specific aptitudes do not provide any greater prediction than *g* alone, then the efficient classification of individuals

in jobs based on specific aptitudes and group aptitudes is not a pertinent issue. However, if there are several factors that differentially predict performance in various jobs, then classification efficiency is a relevant issue. Specific aptitude theory, on the other hand, suggests that job performance is best predicted by one or more specific aptitudes required by the job, rather than by general cognitive ability. For example, a person's performance as an editor would be better predicted by verbal and perceptual speed abilities than by *g* alone. According to this theory, *g* has only an indirect relation to job performance, because it is mediated by specific aptitudes. This theory strongly contributes to the concept of situational specificity to explain differences in job requirements in different settings.

In the latent variable context, the hierarchical model is one of the important cognitive ability theories. Gustafsson (1988) explored the hierarchical model of abilities that considers general and specific factors simultaneously. He proposed a general factor, *g*, that influences the performance of examinees on all tests. In addition, he mentioned a small number of broad factors that influence groups of tests. Hierarchical models of the structure of cognitive abilities offer theoretical as well as practical advantages (Gustafsson, 1988). Such models may resolve the conflict between theorists who emphasize one general ability (e.g., Spearman, Jensen, and Humphreys), and theorists who emphasize several specialized abilities (e.g., Thurstone, Guildford, and Gardner), by allowing for both categories of abilities in the model. The hierarchical approach also offers possibilities for solving prediction problems efficiently and parsimoniously (Gustafsson, 1988; Muthén, 1994).

Recent studies have shown general ability to be a good predictor of future occupational performance (Ree & Earles, 1991; Earles & Ree, 1992; Olea & Ree, 1994). According to the development of the hierarchical ability model, the contribution of specific factors is an important indicator for selection and classification. In the study of general and specific abilities as predictors of school achievement, Gustafsson & Balke (1993) investigated the relationship between aptitude variables and school achievement using a model of ability that allowed simultaneous identification of general and specific abilities. They concluded that differentiation among at least a limited number of broad abilities would be worthwhile. Similar models have been successfully applied to studying subgroup differences in mathematics achievement data (Muthén, Khoo, & Goff, 1994). For most predictive tests, it is practical to include both general and specific factors in the prediction model. For example, to predict a person's performance in a special training program, we would be interested in knowing their ability on the factor related to that particular program and less concerned about their ability on other factors. This type of modeling is useful for finding the relative importance of general and specific factors as predictors of the criterion variable. Muthén and Gustafsson (1995) used this approach to predict job performance. Their study showed that different abilities were of differential important in different jobs.

The purpose of the current study is to investigate the use of factor scores for improving predictive validity in personnel selection and to improve the person-job match. Latent variable modeling will be used to assess predictive validity. Latent variable modeling is useful for identifying the constructs that underlie a set of test items; it makes

more efficient use of the test information than a model which does not use latent variables (Hsu, 1991; Muthén & Hsu, 1993; Hsu, 1995), and it is a useful way of explicating the validity of a test in predicting future performance (Gustafsson, 1988; Hsu, 1991; Muthén & Hsu, 1993; Hsu, 1995; Muthén & Gustafsson, 1995). This report addresses the question: will the approach of using both general and specific factor scores perform better than the method of prediction of using the general factor score only? From a broader perspective, this report contributes to an improvement of the person-job match, which is in the national interest given that it will result in better overall job performance and more effective utilization of the skills of selected individuals.

The current study will use a Monte Carlo approach to assess the prediction of future performance. In addition to the Monte Carlo study, a real data application will be carried out to examine the practical implications of the use of factor scores. The Army Project A and the Marine Corps Job Performance Measurement (JPM) enlistment and performance data have been chosen as an application of the methodology to be considered.

CHAPTER 2

LITERATURE REVIEW

The literature review is organized into four main sections. The first section addresses the issues of factor score determinacy and factor score estimation. The second section reviews the theory and practice in selection and classification. The third section looks at the differential assignment theory (Johnson & Zeidner, 1991). The last section reviews the prediction related to the contribution of *g* and *s* and the concept of cross-validity.

2.1 Factor Score Determinacy and Factor Score Estimation

Beginning with Spearman (1927), analysts believed that the factor score indeterminacy problem was resolvable by simply adding more observed variables in order to better define a factor. Cotter and Raju (1982) showed that the use of factor scores in a regression equation will significantly improve the prediction of population squared cross-validity as compared to the straightforward use of data-level variables. Acito and Anderson (1986) developed a simulation model to investigate the correlations between the true and derived common factor scores under various data conditions. Indeterminacy was found to depend on the level of communality and to be detected more accurately via image factoring than by principal axis or principal component analysis.

Because the theoretical scores are never available, several systems for estimating the factor scores have been proposed. As Tucker (1971) noted, there are four well-known

factor score estimators, as follows: (1) Thurstone proposed regression estimates of the factor scores, a technique which has proven to be very popular; (2) Bartlett proposed estimates that minimized the sum of squared residuals weighted by the reciprocals of the unique variances on the attributes; (3) Horst characterized the least square solution; and (4) Anderson and Rubin proposed a variant of the Bartlett estimates for uncorrelated factors. In addition, Horn (1965) classified several techniques for computing factor scores, which can be described as (a) "exact" procedures that are based upon some form of least squares calculation, and (b) "inexact" procedures in which the factor score is estimated as a simple linear weighted sum of variable scores. Under the system developed by Horn, a procedure is characterized by the type of weighting factor that is used. The property of the factor score estimators has been examined with respect to internal and external characteristics. Internal characteristics reflect the covariation of the estimates of other factors and with the theoretical factors. External characteristics reflect the relationship between the factor scores and the variables outside the factored battery.

Despite the disagreements in the literature about factor scores, and the availability of several methods for constructing scores, few studies have examined indeterminacy directly by means of empirical or simulated data. Wackwitz and Horn (1971) started with "known" factor scores and generated "observed data" using a Monte Carlo simulation to compare principal component scores with common factor scores derived from two factoring methods. Because there are several procedures for estimating factor scores, the question of which factor score estimation procedure is better for the purpose of prediction has received a major share of the attention in studies conducted by Morris (1979, 1980).

Morris' research was based on Monte Carlo simulation techniques, and there is still a great need for studies comparing different factor score estimates using empirical data. Without empirical studies, the decision-makers may continue to fail to realize the benefits of factor scores in prediction. The few empirical and simulated data comparisons reviewed leave many unanswered questions. Empirical studies of analytic procedures using real data have their place, but their limitation is that the "correct" answer is not known. In contrast, simulation studies can start with well-defined, known parameters, thereby allowing one to assess the relative accuracy of techniques under different data conditions.

Short (1990) found that determinacy for the general factor was robust in the following sense: regardless of the size of the factor loadings, the number of items in the model, or the number of items influenced by specific factors, the general factor determinacy remained high. Even when specific factor loadings were higher than those for the general factor, the scores were very reliable. The size of the specific factor loading was a major influence on the specific factor determinacy, as was the size of the general factor loading, although to a lesser extent. The portion of items influenced by specific factors was found to be more influential in obtaining reliable factor scores than the number of items overall. Therefore, the specific factor scores were not as reliable as expected. Short's study also demonstrated that specific factor scores from multidimensional, dichotomized data are not extremely reliable. It was suggested that the initial test or attitude data should be continuous and have at least moderate factor loadings in order to create reliable scores.

Research in the area of selection and predictive validity with latent variable structures (Hsu, 1991; Muthén & Hsu, 1993; Hsu, 1995) found that the factor score estimator performed as well as the theoretical optimal method, which is a full quasi-likelihood estimator (FQL). When regressing criterion on the estimates of factor scores in a random sample, it gave consistent estimates of regression slopes. Although the selected sample was not a random sample, when selection was based on the latent factors, or on the estimated factor scores, the estimated regression slopes were still unbiased (Muthén & Jōreskog, 1983).

2.2 Selection and Classification

The general framework of problems in prediction involves two elements: unknown criterion (or future performance) and known predictor. The desire is to express uncertainty about criteria in the light of predictors. For instance, we typically appeal to a model that formalizes judgment about how predictor and criterion are related. If selection was based on the factor scores, the regression of future performance on factor scores would not be distorted (Muthén & Hsu 1993). A major advantage of using factor scores is because the estimated factor score is a linear transformation of observed variables, the factor score method produces unbiased estimates of slope when selection is based on observed variables.

Most research in prediction has a similar observed data pattern with missing data on criterion variable as shown in Figure 1. This figure shows 10 subtests for each

applicant ($i = 1$ to N) while the criterion is observed only for the selected group ($i = 1$ to N_S). For the non-selected group ($i = N_{S+1}$ to N), there is no information for the criterion. The information from the criterion variables is collected only for those being selected. If the selection procedure is based on the observed predictors, the data can be said to be missing at random (Little & Rubin, 1987). Assuming missing at random, maximum-likelihood estimation of latent variable models with missing data can be carried out by structural equation modeling techniques (Muthén, Kaplan, & Hollis, 1987).

In the latent variable approach, the criterion variable is regressed on the factors. The fact that information on the predictors is used from both the selected and non-selected groups adjusts for the selectivity in the selected group. This approach is analogous to the conventional Pearson-Lawley adjustment. Muthén and Hsu (1993) and Hsu (1995) showed that factor score estimates perform about as well as maximum-likelihood estimates with respect to estimated regression coefficients, standard errors, standardized coefficients, and R^2.

Research in job classification has focused on the appropriate data analysis model for analyzing the similarities and differences among jobs. In the research performed by Cornelius, Carron, and Collins (1979), the data analysis model was held constant and the type of job analysis data was varied to examine the effect on the resulting job classification decisions. It is important to realize that jobs can be similar and dissimilar among different levels of analysis. Cornelius et al. (1979) suggested that the selection of the appropriate job analysis model is at least as important as the selection of the appropriate data analysis model in job classification research. Also, Cornelius, Schmidt,

and Carron (1984) raised three different ways in which job analysis is important for selection purposes: (1) determining whether jobs are similar enough to be combined into a single selection system; (2) identifying knowledge, skills, and abilities (or aptitudes) that are important for job performance; and (3) determining whether a test can be transported from a setting in which it has been shown to be valid to a new setting.

Selection tests are used to accept or reject an applicant for a job. Once an organization accepts an individual for employment, classification tests may then be used to assign an individual to a specific training program of a job from among a number of available opportunities. Meanwhile, the purpose of classification is to match individuals and jobs in a manner that maximizes aggregate performance. Classification decisions are a major concern in military services and are becoming an of increasing interest to industry. Classification is also used in counseling in order to provide guidance to students in the choice of a field of study or an occupation. Furthermore, classification is utilized in clinical diagnosis in order to aid in the choice of a course of treatment (Zeidner & Johnson, 1991b). Traditionally, in selection and placement decisions, only a single job is involved and can be accomplished with one or more predictors. The outcome of selection and placement is determined by an individual's predicted performance. Classification, however, requires multiple predictors measuring more than one dimension of job performance. Alley (1994) noted that the concept of classification was broadly defined to include selection as a special case. If there are multiple vacancies for one job category and the number of applicants is greater than the number of job vacancies,

selection will occur. Classification usually implies multiple job categories and may or may not involve some number of nonselectees.

Vineberg and Joyner (1989) made the standard distinction between job proficiency and job performance: namely, contrasting what a person knows or can do with what a person actually does on the job. Proficiency usually is measured by a paper-and-pencil or a hands-on test of job tasks, and is generally objective and reliable. Job performance measures, usually ratings, are generally subjective and less reliable than proficiency measures. Correlation between written job-knowledge measures of proficiency and hands-on job sample measures of proficiency were generally found to be low, ranging from $r = 0.00$ to about $r = 0.30$. However, when job-knowledge tests were constructed, based only on information directly relevant to job performance, higher correlations were found, ranging from $r = 0.58$ to $r = 0.78$ (Vineberg & Joyner, 1989). The low reliability of ratings limited their relationship with other proficiency measures, with only a few correlations appearing above $r = 0.30$.

2.3 Differential Assignment Theory (DAT)

The underlying thought of the differential assignment theory (DAT) approach is based on the concepts of classification efficiency and differential validity introduced by both Brogden and Horst (Johnson & Zeidner, 1991; Zeidner & Johnson, 1991a, 1991b, 1991c; Scholarios, Johnson, & Zeidner, 1994). The concept of DAT is derived from an integrative review of personnel classification literature, especially the contributions of

Brogden and Horst, combined with the systematic development of methodologies for improving classification efficiency.

Zeidner and Johnson (1991a) proposed DAT, postulating that several factors differentially predict performance in various jobs. They believe that DAT provides a more coherent framework for job classification, while still recognizing *g* as the dominant predictor of performance. DAT stresses the difference between predicted performance measures across jobs and explains classification efficiency as a function of mean predictive validity, mean intercorrelation among predicted performance measures, the number of jobs to which individuals are assigned, and the selection ratio. DAT states that the joint predictor-criterion space is multidimensional with useful factors contributing a nontrivial amount of classification efficiency in addition to the unidimensional space defined by the *g* factor (Zeidner & Johnson, 1991a).

There seems to be ample value in tests of general mental ability and tests of specific abilities. The *g* proponents argue that the best way to classify large numbers of applicants in terms of probable success is with a measure of general intelligence. Other research also suggests a continued interest in ability measurement beyond an exclusive psychometric *g* approach. There is good reason to believe that the relevant issue is not whether psychometric *g* or a measure of a specific ability is a better predictor of job success, but rather what are the limiting conditions to the use of either approach. Lohman (1994) has applied the more orthodox cognitive science paradigm to the skilled performance problem and has come to the same conclusion with respect to the value of a multi-ability view of the world. The classic model says simply that the general factor

will account for almost all the relevant true score covariances among observed measures. The goal of measurement is to obtain the best possible measure of the general factor. The multiple-factor model assumes that performance is multidimensional and is composed of a number of basic distinguishable components, which are such that some people could perform well on one component but not as well on others.

Scholarios, Johnson, and Zeidner (1994) illustrated that "Selection into the organization is first accomplished with a single composite resembling a measure of general cognitive ability (*g*); assignment is then made to specific jobs with weighted test composites tailored for each job" (p. 412). Classification efficiency was measured as mean predicted performance determined after optimal assignment to jobs. Their study provided a comparison of differential assignment theory with general aptitude theory and validity generalization. The results provided evidence that efficient classification, using tailored (best weighted) test composites to optimally assign new soldiers to a set of jobs, is best accomplished by the design of a test battery of multidimensionality in the predicted performance space. The theoretical value and practical usefulness of DAT was supported by the finding that both longer test batteries and the use of Horst's differential validity index to select tests increase potential classification efficiency.

2.4 Predictive Validity and Cross-Validity

2.4.1 Predictive Validity in Relation to g and s

In previous studies (Ree & Earles, 1991, 1992; Ree, Earkesm, & Teachout, 1994), when general (*g*) and specific (*s*) abilities were used to predict training grades and job

performance, it was found that g was the most potent predictor and that s added little to prediction. Ree, Earkesm, & Teachouts' study (1994) extended the finding of statistically significant but practically small incremental validity for specific measures to seven additional jobs and to new criteria. It also showed that the incremental value of the specific measures was small for all three criteria and demonstrated the application of estimates of effective sample size in the computation of adjusted multiple correlation coefficients. The average increment to g by measures of s was 0.21, about the same as was found in previous studies for both training criteria (Ree & Earles, 1991) and job performance criteria (McHenry et al., 1990). It is also consistent with the estimate provided by Hunter and Hunter (1984). Carey et al. (1994) studied the predictive efficiency of adding new tests to a highly g-saturated test battery for the prediction of both job performance and training criteria; he found increments averaging 0.02 across these criteria. Morales and Ree (1992) found similar incremental differences for predicting pilot and navigator performance that included work sample criteria. The results showed that g was the best predictor of pilot and navigator job performance in a study of 5,500 airmen. The average validity of g was 0.33 and the average increase from non-g was 0.05.

Prediger (1989) challenged the conclusion (Hunter 1986; Jensen 1986; Thorndike 1986) that general ability is more important in determining occupational level and job performance than specific abilities, but he presented no data on validity or incremental validity. His conclusion that specific aptitudes are important in performance was based on data showing distinct patterns of specific-aptitude means across occupations, both

among incumbents and among high school students who later entered specific occupations. However, he stated that validities can be equal for different occupations when means are different.

Hunter (1985) reviewed meta-analyses of hundreds of studies showing that general cognitive ability predicts job performance in all jobs, whether performance is measured objectively or subjectively. He also reviewed path analytic research consistent with the theory that *g* affects job performance primarily by improving job knowledge, but that general ability also affects job performance above and beyond its impact on job knowledge. Hunter discussed evidence that, except in a few special cases, tailoring aptitude composites to match the job does not improve the prediction of job performance above and beyond that provided by general cognitive ability.

In the Army Project A, three types of measurement were used: (1) hands-on tests of job performance, (2) multiple-choice tests of job knowledge, and (3) ratings of job performance. Project A was designed to focus on individual differences in predictors and on performance measures, and to evaluate the relationship between predictors and criteria for a wide variety of very different individuals (Green & Wing, 1988). The Army researchers found that each job was composed of both elements unique to that job and elements shared by all jobs in the Army. The findings also showed that all predictors were not equally valid for the different aspects of job performance. Wise & McHenry (1990) also concluded that job performance is multidimensional. Their major findings were that different predictor equations were needed for each of the five criterion factors.

In addition, different prediction equations were required for the component that reflected proficiency in the technical tasks specific to each job.

2.4.2 Cross-Validity

The predictive power of a sample regression equation in the population and in future samples is often of primary importance to researchers. A measure widely used for this purpose is the squared cross-validity coefficient, R_c^2. This index is defined as the squared correlation of actual criterion values with those predicted from the sample equation for the population of interest. A natural choice as an estimator of this parameter is the sample squared multiple correlation (Kennedy, 1988).

Most authors suggest splitting sample data, then using one portion for identification of the model and the other portion for estimation of parameters. But cross-validation is known to have significant restrictions. In particular, a significant loss of information can be expected when all available data are not used for purpose of parameter estimation. When sample size is large, this loss in most likely minor. But for a moderate size dataset, splitting data can yield seriously unstable parameter estimates. Previous research has shown that sample size and the ratio of predictors selected to the total in the set will affect validity estimation in the subset context.

Morris and Guertin (1977) showed the superiority of factor scores as measured by the cross-validity correlations. They compared common factor scores to unfactored data-level variables as predictors in a regression equation. The Monte Carlo study by Morris and Guertin showed that the regression equations using factor scores resulted in less shrinkage as a result of cross-validation than the data-level variables in all cases.

Cotter and Raju (1982) conducted a study to evaluate formula-based population squared cross-validity and estimates of factor scores in prediction. They concluded that formula-based estimates of population squared cross-validity are as good as those obtained from the conventional cross-validation procedure.

There are four correlations that are important in a validation study involving several predictors: the sample square multiple correlation (r^2), the sample squared cross-validity correlation (r_c^2), the population squared multiple correlation (ρ^2), and the population squared cross-validity correlation (ρ_c^2). The population squared multiple correlation is the square of the multiple correlation developed on the entire population. The population squared cross-validity is based on the regression weights developed on a sample applied to the entire population. The most important correlation in selection is the population squared cross-validity, since it provides a measure of how well a regression equation developed on a sample will do in future sampling from the population. However, in most situations, the population of interest is not available and the population squared cross-validity correlation cannot be calculated directly. Consequently, the sample squared cross-validity is viewed as an estimate of population squared cross-validity. The cross-validation procedure does have a serious drawback in practice. If the original sample is small, the splitting of the sample into two subsamples of even smaller size is known to effect the stability of the regression equation, thus raising questions about the practicality of the cross-validation procedure. Cotter and Rajus' study (1982) suggested that the formula-based estimation of population squared cross-validity is

satisfactory, and there is no advantage in conducting a separate, expensive, and time consuming cross-validation study.

CHAPTER 3

METHODOLOGY

The present study aims to investigate whether the approach of prediction by using specific factors in addition to the general factor will perform better than the method of prediction of using the general factor score only. Technical definitions of predictive validity and factor score determinacy in the context of a latent variable model and the definition of sensitivity, specificity, and proportion of correct decisions in the context of the decision table are discussed and defined. In the next section, the Spearman rank-order correlation coefficient and the cross-validation are described as indicators of the stability of the prediction used in the selection and classification. The final section describes the design of the Monte Carlo study used to explore the use of factor scores for prediction.

3.1 Technical Definitions

3.1.1 Measurement Model

The measurement model (factor analysis) is written as

$$x = \nu + \Lambda\eta + \varepsilon,$$

where x contains the predictor variable and ε is the residual, with covariance matrix Θ. The matrix Λ consists of factor loadings and the vector η consists of factor variables. We have

$$E(\varepsilon) = 0, \qquad Var(\varepsilon) = \Theta \text{ diagonal}.$$

We assume

$$E(\eta) = \alpha, \qquad Var(\eta) = \Psi.$$

Then

$$Cov(x) = \Sigma_x = \Lambda\Psi\Lambda' + \Theta.$$

Note that the off-diagonal elements of Σ_x are functions of Λ and Ψ; they do not involve Θ.

For identification we may require

$$\Psi = I.$$

This model is known as the orthogonal factor analysis model. Then $\Lambda\Psi\Lambda' = \Lambda\Lambda'$, which leaves the indeterminacy of the multiplication of Λ on the right by an arbitrary orthogonal matrix.

3.1.2 Predictive Validity

Consider next the prediction equation,

$$y = \alpha + \beta'\eta + \delta,$$

where the criterion variable y is regressed on the factor scores (η). β is the regression slope of y on η. The factor scores are estimated by using the information on the predictor variables (x) from the applicant sample. It is known that for a random sample, regressing y on the regression estimates of factor scores gives consistent estimates of regression slopes (Tucker, 1971; Hsu, 1991; Muthén & Hsu, 1993; Hsu, 1995). Even if the sample is not a random sample, when selection is based on the estimated factor scores, the regression slope of y on the factor scores is still unbiased. In this report, predictive

validity is defined as the regression coefficients of the criterion variable y on the latent factor η.

3.1.3 Regression Method of Factor Score Estimation

Factor scores may be estimated for each latent variable in the factor analysis framework. For continuous variable x, the standard regression method gives the estimates

$$\eta = \Psi\Lambda' (\Lambda\Psi\Lambda' + \Theta)^{-1} x$$

when both observed and latent variables are standardized to zero mean (Harris, 1967; Maxwell, 1971; Tucker, 1971; Muthén, 1978).

The procedures for the regression method of factor score estimation are: (1) confirmatory factor analysis on the applicant sample by the LISREL8 program in order to determine the factor loadings (Λ) and the variances of factors (Ψ); (2) the regression method for computing the estimated factor scores by SAS IML using the factor model estimates; and (3) regression of the criterion variable on the estimated factor scores for the selected sample.

3.1.4 Factor Score Determinacy

Because the theoretical factor scores are never available, it is important to investigate the correlations between the true and the estimated factor scores. These values are the factor score determinacies. When the ηs are standardized, Ψ is equal to I. In this case, determinacy may be seen to depend on the matrix of factor loadings (Λ),

$$E(\eta, \hat{\eta}) = \Omega = \Lambda' [\Lambda\Lambda' + (I - \mathrm{diag}(\Lambda\Lambda'))]^{-1} \Lambda.$$

The correlations between the true score (η) and the estimated factor scores ($\hat{\eta}$) are the square roots of the diagonal elements of Ω (Maxwell, 1971; Short, 1990).

3.1.5 Sensitivity, Specificity, and Proportion of Correct Decisions

A sensitivity analysis is used to evaluate how the prediction (or selection decision) is affected by using the general factor score only to predict the criterion compared to using the specific factor scores in addition to the general factor score. Allred (1991) illustrated that a 2 × 2 decision table provides a method for evaluating the cost or utility of a test score cutoff. Once the selection ratio or the cutoff has been determined, a simple 2 × 2 table can be used to display the number of successes and failures in the selected and rejected groups. The general form of the 2 × 2 table is shown in Figure 2. With this method, it is simple to determine the number of correct decisions made about individuals. Accepting an individual who succeeds and rejecting an individual who would fail are correct decisions. The proportion of correct decisions is the total of correct accept (TP) and correct reject (TN) divided by the total number of individuals (N). In Figure 2, three equations for computing sensitivity, specificity, and proportion of correct decisions are shown. Sensitivity is defined as the proportion of successful individuals who are accepted, while specificity is the proportion of failing individuals who are rejected.

From the economic perspective of the employer, the worse error is to hire a poor worker. Therefore, the sensitivity analysis is sufficient for cost-efficient recruiting.

3.1.6 The Spearman Rank-Order Correlation Coefficient

The Spearman rank-order correlation coefficient is used to correlate two continuous variables that are measured at the ordinal level of measurement. If the subjects are rank-ordered (from highest to lowest) on each of the two variables and the ranks are correlated using the Pearson correlation coefficient, the resulting number is a Spearman rank-order correlation coefficient. The Spearman rank-order correlation coefficient is used to evaluate the rank-ordered on the estimated criterion using different methods. The correlation coefficient will be computed for the ranking of the criterion and the ranking of the estimated criterion.

3.1.7 Cross-Validation

In cross-validation, the data are split into two or more subsets. One of the subsets is called the construction set, and is used for estimation. Predictions for the cases in the other subset, called the validation set, can be obtained from the model fit to the construction set using predictor values from the validation set. These predictions can be compared to the observed values of the response.

However, cross-validation is known to have significant restrictions. In particular, a significant loss of information can be expected when all available data are not used for the purpose of parameter estimation. When the sample size is large, this loss is most likely minor. But for a moderate-size dataset, splitting data can yield seriously unstable parameter estimates. Previous research has shown that sample size and the ratio of predictors selected to the total in the set will affect validity estimation in the subset context. One useful criterion function to determine the outcome of the cross-validation is

the square root of the average square error of prediction (PRESS); thus, a good model indicates small value of PRESS (Weisberg, 1985). It is expected that the value of PRESS for the method of using both general and specific factors is smaller than that for the method of using the general factor only.

For the simulation study as shown in Figure 3, the selected sample (N_S) is split into two subsets (e.g., sample 1 and sample 2). The construction set (sample 1) is used for identification of the model (estimation of regression slopes) and the validation set (sample 2) is used for estimation of criterion. For the validation set, PRESS and Spearman rank-order correlation are calculated according to the hierarchical factor model ($\hat{g} + \hat{s}$) and the single-factor model ($\hat{g}_I$). Double cross-validation is carried out in this report so the procedure outlined above is applied twice (sample 2 is used for estimation and sample 1 is used for prediction). So, for each sample, the regression equation and the predicted criterion are calculated. If the results of double cross-validation are close, as suggested by Pedhazur (1982), we may combine the samples and calculate the regression equation to be used in prediction. The purpose of this method is to study the differences of Spearman rank-order correlation coefficients as well as the differences of PRESS across two models in the context of double cross-validation.

3.2 The Monte Carlo Study

This report aims to examine the quality of factor scores as predictors to improve the precision of selection and classification. The classic model says that the general

factor will account for almost all the relevant true score covariances among observed measures. It uses only general ability as a predictor of future performance or as a selection criterion to screen applicants. On the other hand, the multiple-factor model assumes that performance is multidimensional and is composed of a number of basic components, which are such that some people may perform well on one component but not so well on others. In this report the specific factors, in addition to the general factor, are included as predictors. The regression method is used to estimate factor scores because it is a linear transformation of observed variables. The selection is based on factor scores so as to give unbiased estimates of regression slopes. The research question is: Does the method of prediction of using general and specific factor scores perform better than the method of prediction of using the general factor score only?

A simulation model is developed to explore the use of factor scores in prediction under conditions varying the following: (1) the factor loadings for the specific factor in terms of high vs. low determinacy; (2) the regression slopes for the criterion regressed on general and specific factors; (3) the selection methods; (4) the R square; (5) the selection ratio; and (6) the sample size for the selected subjects. Next, the prediction of future performance is examined in terms of sensitivity, specificity, and proportion of correct decisions. In addition, two approaches, cross-validation and Spearman rank-order correlation, are used to evaluate the accuracy of prediction.

The Monte Carlo study is intended to answer the following questions: How does the quality of factor scores affect the selection and prediction? Does the inclusion of the specific factor enhance the predictive validity? Does the selection based on the specific

factor in addition to the general factor increase sensitivity, specificity, proportion of correct decisions, and/or rank correlation? How reliable must factor scores be in order to improve predictions? How do the R square, the regression slope, the selection ratio, and the sample size affect the predictive validity?

3.2.1 Design of the Monte Carlo Study

The design of this Monte Carlo study is summarized in Table 1. This report considers 4 datasets (2 selection ratios × 2 sample sizes for the selected individuals) and 36 latent variable models (12 models × 3 cases), resulting in 144 combinations. The three cases are named as $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$, where "g" represents the general factor and "s" represents the specific factor. The notation of $g + 3\ s$ means one general factor plus three specific factors (s_1, s_2, and s_3). For each of the 144 conditions, one hundred replications are performed.

Table 2 shows the parameter values for each model across the three different cases. The standardized factor loadings, variance of factors and criterion, regression slopes, standardized regression slopes, and R square are presented.

Selected Sample Size and Selection Ratio

In order to obtain the actual prediction situations, two different sample sizes for the selected individuals are studied. Selection ratios of 0.50 and 0.10 are used in the simulation. The sample sizes for the selected subjects are 250 and 500, corresponding to the applicant sample sizes of 500 and 1,000 with a 50% selection ratio, and the applicant sample sizes of 2,500 and 5,000 with a 10% selection ratio.

Selection Methods

In the simulated data, the criterion variable is available for all applicants so the true selectee is determined by the observed criterion (y) and the selection ratio. Two selection methods are used: the estimated general factor score ($\hat{g}_1$), and the predicted criterion ($\hat{y}$) that is estimated by general and specific factor scores ($\hat{g}$ and $\hat{s}$). These factor scores are computed using the regression method for all applicants. Given that previous research (e.g., Ree & Earles, 1991; Earles & Ree, 1992; Ree, Earles, & Teachout, 1994; Ree & Carretta, 1995) uses the general ability as selection criterion, the current study will use the same selection criterion as the standard for comparison. The classification of applicants as accepted or rejected is based only on the ranking of the estimated general factor score. In contrast, the decision method used in this report is based on the predicted criterion ($\hat{y}$) by using both general and specific factor scores as predictors. The purpose is to compare the efficiency of these two selection methods.

Two types of latent variable models as shown in Figures 4 and 5 will be considered as examples for the $g + 1\ s$ case. The first type is a single-factor model that has only one general factor (g_1). The second type is a hierarchical factor model that has a general factor (g) which influences all of the subtests (x_1 to x_{10}) as well as one specific factor (s) which influences some of the subtests (x_6 to x_{10}). In Figures 4 and 5, x_1 to x_{10} are the observed variables and y is the observed criterion variable. For the single-factor model, the criterion variable is influenced by one general factor (g_1), which represents general ability. For the hierarchical factor model, the criterion variable is influenced by

one general factor (g), which represents general ability, and one specific factor (s), which represents a special, narrow ability. The factors are all uncorrelated with each other. Since the hierarchical factor models for $g + 2\ s$ and $g + 3\ s$ are similar to that for $g + 1\ s$, they are not presented here.

True Models

Table 1 shows the 36 model combinations. Models 1 through 6 are used to differentiate the effects of predictive validity for g and s and also specific factor determinacy, while controlling for the value of R^2 (0.4) and general factor loadings. Models 7 through 12 have the same pattern as described in models 1 through 6, but the value of R^2 is set at 0.6.

The range of g factor loadings is from 0.25 to 0.80 for all 12 models in each of the three cases. The range of s factor loadings for the case of $g + 1\ s$ is from 0.18 to 0.54 in models 1, 3, 5, 7, 9, and 11, and from 0.36 to 0.72 for models 2, 4, 6, 8, 10, and 12. In models 1 and 2, the standardized regression slope for g (0.54) is much higher than that for s (0.32) and the variances of g and s are set at 1.00 and 0.80, respectively, while in models 3 and 4, standardized regression slopes for g (0.45) and s (0.45) are nearly equivalent and the variances of g and s are set at 1.00 and 0.81, respectively. In models 5 and 6, the standardized regression slope for g (0.32) is less than that for s (0.55) and the variances of g and s are set at 1.00 and 0.81, respectively. Models 7 through 12 have the same pattern of regression slopes as described for models 1 through 6.

For the $g + 2\ s$ case, the range of low specific factor loadings is from 0.13 (0.12) to 0.54; and the range of high specific factor loadings is from 0.32 (0.30) to 0.72. In

models 1 and 2, the standardized regression slope for g (0.49) is much higher than that for s_1 (0.33) and s_2 (0.23) and the variances of g, s_1, and s_2 are set at 1.00, 0.80, and 0.40, respectively, while in models 3 and 4, standardized regression slopes for g (0.36), s_1 (0.36), and s_2 (0.37) are nearly equivalent and the variances of g, s_1, and s_2 are set at 1.00, 0.81, and 0.36, respectively. Furthermore, in models 5 and 6, the standardized regression slope for g (0.32) is less than that for s_1 (0.39) and s_2 (0.39) and the variances of g, s_1, and s_2 are set to 1.00, 0.81, and 0.36, respectively. The parameters for models 7 through 12 follow the same pattern described in models 1 through 6.

For the $g + 3\ s$ case, the two sets of range for specific factor loadings are from 0.23 (0.24) to 0.45 and from 0.23 (0.24) to 0.36. In models 1 and 2, the standardized regression slope for g (0.43) is much higher than that for s_1 (0.38) , s_2 (0.20), and s_3 (0.16) and the variances of g, s_1, s_2, and s_3 are set at 1.00, 0.80, 0.60, and 0.40, respectively; in models 3 and 4, standardized regression slopes for g (0.32), s_1 (0.32), s_2 (0.31), and s_3 (0.32) are nearly equivalent and the variances of g, s_1, s_2, and s_3 are set at 1.00, 0.81, 0.64, and 0.36, respectively; in models 5 and 6 standardized regression slope for g (0.31) is less than that for s_1 (0.37), s_2 (0.49), and s_3 (0.37) and the variances of g, s_1, s_2, and s_3 are set at 1.00, 0.81, 0.64, and 0.36, respectively . Models 5 through 12 have a combination of regression slopes and factor variances similar to that seen in models 1 through 6.

Factor Determinacy

It is clear that increasing the number of observed variables increases the reliability of the factor score measurement (Acito & Anderson, 1986; Short, 1990). In this

simulation, the effects on specific factor determinacies are investigated by varying the factor loadings. When the general factor loadings are held constant, increasing the specific factor loadings should increase the reliability of the specific factor. Two sets of standardized factor loadings for specific factors (low λ_s vs. high λ_s) are used to investigate the issue of determinacy.

As expected, increasing the size of loadings increases the factor determinacy. The determinacy gives information about the reliability of the factor. In the current study, the general factor determinacy is about 0.94 for the case of $g + 1\ s$, from 0.90 to 0.92 for the case of $g + 2\ s$, and from 0.90 to 0.91 for the case of $g + 3\ s$ when general factor loadings are from 0.25 to 0.80.

In the $g + 1\ s$ case, for models 2, 4, 6, 8, 10, and 12 with high specific factor loadings (ranging from 0.50 to 0.80), the specific factor determinacy is about 0.85. In contrast, for the odd-numbered models with low specific factor loadings (ranging from 0.20 to 0.60), the specific factor determinacy is about 0.70.

In the case of $g + 2\ s$, models with high specific factor loadings (ranging from 0.50 to 0.80) have the specific factor determinacies of about 0.84 and 0.71 for models 2 and 8; and 0.85 and 0.68 for models 4, 6, 10, and 12. In contrast, models with low specific factor loadings (ranging from 0.20 to 0.60) have the specific factor determinacies of about 0.68 and 0.53 for models 1 and 7; and 0.68 and 0.51 for models 3, 5, 9, and 11.

In the case of $g + 3\ s$, for models 2, 4, 6, 8 10, and 12 with high specific factor loadings (ranging from 0.30 to 0.80), the specific factor determinacies are about 0.79,

0.76, and 0.75. In contrast, for models 1, 3, 5, 7, 9, and 11 with low specific factor loadings (ranging from 0.30 to 0.70), the specific factor determinacies are about 0.62, 0.72, and 0.61.

3.2.2 Analyses

The following steps are carried out in the Monte Carlo study:

(1) generation of observed variables (x_1 to x_{10}), criterion variable (y), and true factor scores on general (g) and specific (s) factors for N applicants as shown in Figure 6;

(2) estimation of factor score determinacies;

(3) estimation of parameters (Ψ and Λ) based on the single-factor model and the hierarchical factor model;

(4) estimation of factor scores ($\hat{g}_I$, $\hat{g}$, and $\hat{s}$) for each individual;

(5) obtaining a selected sample (N_S) according to the estimated general factor scores ($\hat{g}_I$) and the selection ratio;

(6) estimation of regression slopes ($\hat{\beta}_g$ and $\hat{\beta}_s$) for general and specific factors with the criterion using the selectees N_S from the previous step;

(7) estimation of the predicted criterion ($\hat{y}$) using the estimated regression slopes and the estimated factor scores (this new selection method is defined as the equation below:

$\hat{y} = \hat{\alpha} + \hat{\beta}_g \times \hat{g} + \hat{\beta}_s \times \hat{s}$);

(8) obtaining a new selected sample based on $\hat{y}$ and the selection ratio;

(9) categorizing applicants as success or failure based on the true criterion y and selection ratio ;

(10) computing sensitivity and specificity for the classifications of the true criterion (y) and the predicted criteria ($\hat{y}_{\hat{g}_1}$ vs. $\hat{y}_{\hat{g}+\hat{s}}$);

(11) computing the Spearman rank-order correlation coefficient between the rank-ordered y and the rank-ordered $\hat{y}_{\hat{g}_1}$;

(12) computing the Spearman rank-order correlation coefficient between the rank-ordered y and the rank-ordered $\hat{y}_{\hat{g}+\hat{s}}$;

(13) double cross-validation:

a) splitting of the selected sample N_S into two subsets (labeled as sample 1 and sample 2), the construction set for estimation and the validation set for prediction (see Figure 3),

b) estimation of the criterion function, the square root of the average squared error of prediction, the equation of PRESS is defined in Figure 3,

c) computing of the Spearman rank-order correlation coefficient between the rank-ordered y and the rank-ordered $\hat{y}$,

d) repeating steps a through c, where sample 2 is for estimation and sample 1 is for prediction.

CHAPTER 4

RESULTS OF THE MONTE CARLO STUDY

In this chapter, the results of the Monte Carlo study are presented. The purpose of the Monte Carlo study is to examine the quality of factor scores as predictors to improve the precision of selection and classification of applicants. First, the results of classification and Spearman rank-order correlation coefficient are described and summarized. Next, the results are described as a function of selection methods, selection ratio, and sample size. The effects of model features in terms of R square, standardized regression slope, and factor loadings are then presented. The last section shows the increment of prediction from the specific factors.

4.1 Results of Sensitivity, Specificity, and Proportion of Correct Decisions

The analysis results are summarized in Tables 3 through 5 for $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$, where the classifications in different combinations of sample sizes for the selected individuals, selection ratios, models, and selection methods are shown. The pattern of results is virtually identical for all of the models, so only one set of results is shown for $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$.

g + 1 s

Model 1 in the first section, defined as N=250, R=0.5 in Table 3, shows the 50% selection ratio (R=0.5) in the sample of 500 applicants (250/0.5=500). The first block of

$\hat{g}_1$ presents selection results using the method in which the predictive criterion is based on the estimated general factor score ($\hat{g}_1$). The results show that 68.75% of the successful subjects were also selected by the predictive criterion using $\hat{g}_1$, but that 31.25% (1 - 68.75%) of the successful subjects were excluded by this criterion. The results show that about 156 subjects (500 × 31.25%), or 31.25% of the applicants, are misclassified by the traditional method ($\hat{g}_1$). In contrast, the second block of $\hat{g} + 1\ \hat{s}$ in Table 3, using both estimated general and specific factors ($\hat{g} + 1\ \hat{s}$), shows that 69.62% of the successful subjects are selected according to the predictive criterion ($\hat{y}$). Of the successful subjects, 30.38% are not selected by the new criterion. The results show that about 150 subjects, or 30.38% of the applicants, are misclassified by the new method. The third block of difference shows the increase in sensitivity, specificity, and proportion of correct decisions that occurs when using specific in addition to general factors as predictors. It shows that there is an increase of 0.87% for sensitivity, specificity, and proportion of correct decisions in model 1 with a 50% selection ratio and sample size 250 for the selected subjects.

g + 2 s

Model 8 in the second section, defined as N=250, R=0.1 in Table 4, displays the 10% selection ratio (R=0.1) in the sample of 2,500 applicants (250/0.1). The results show that 47.88% of the successful subjects were also selected by the predictive criterion using $\hat{g}_1$, but 52.12% of the successful subjects were excluded by this criterion. There are about 260 subjects, or 10.42% (1 - 98.589%) of the applicants, who are misclassified

by the $\hat{g}_1$ method. In contrast, using both estimated general and specific factors ($\hat{g}$, $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_3$) shows that 50.39% of the successful subjects are chosen according to the predictive criterion ($\hat{y}$). Of the successful subjects, 49.61% are not selected by the new criterion. There are about 248 subjects, or 9.92% of the applicants, who are misclassified by the $\hat{g}$ + 3 $\hat{s}$ method. Furthermore, there is an increase of 2.50% for sensitivity, an increase of 0.28% for specificity, and an increase of 0.50% for the proportion of correct decisions.

g + 3 s

Model 12 in the fourth section, defined as N=500, R=0.1 in Table 5, shows the 10% selection ratio (R=0.1) in the sample of 5,000 applicants (500/0.1). The first block of $\hat{g}_1$ presents selection results using the method in which the predictive criterion is based on the estimated general factor score ($\hat{g}_1$). The results show that 34.97% of the successful subjects were also selected by the predictive criterion using $\hat{g}_1$, but 65.03% of the successful subjects were excluded by this criterion. Because the criterion information is observed for all applicants, the results show that about 650 subjects or 13.01% (1 - 86.99%) of the applicants are misclassified by $\hat{g}_1$ method. In contrast, the second block of $\hat{g}$ + 3 $\hat{s}$ in Table 5, using both estimated general and specific factors ($\hat{g}$, $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_3$), shows that 44.39% of the successful subjects are selected according to the predictive criterion ($\hat{y}$). Of the successful subjects, 55.61% are not selected by the new criterion. The results show that about 556 subjects, or 11.12% of the applicants, are misclassified by the $\hat{g}$ + 3 $\hat{s}$ method. The third block shows that there is an increase of 9.42% for

sensitivity, an increase of 1.05% for specificity, and an increase of 1.88% for the proportion of correct decisions in model 12 with a 10% selection ratio and sample size 500 for the selected subjects.

Figures 7 through 9 present the plots of sensitivity, specificity, and proportion of correct decisions across four combinations of sample sizes for the selectees and selection ratios for $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$, respectively. Each plot shows two selection methods and displays by ordering the new method (general plus specific factors). Figures 10 through 12 show the difference in sensitivity, specificity, and proportion of correct decisions between the method of using the general factor only as predictor and the method of using the general and specific factors as predictors for $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$ cases, respectively.

4.2 Results of Spearman Rank-Order Correlations

Spearman rank-order correlations are obtained from the total applicant sample and the cross-validation sample.

4.2.1 Applicant Sample

The Spearman rank-order correlation coefficients between the rank-ordered true criterion and the rank-ordered estimated criterion are given in Tables 3 through 5 for the cases of $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$, respectively. These coefficients are computed for all applicants. Figures 13 through 15 present the plots in ascendant order for the new approach. The results agree with the analyses of sensitivity, specificity, and proportion of correct decisions. It is observed that the pattern of results for rank correlation is similar

to that for classification. Moreover, the results are not affected by selection ratio and sample sizes for the selected subjects. However, the method of using the specific factor in addition to the general factor to predict the rank-ordered criterion performs better than the method of using the general factor alone to predict the rank-ordered criterion.

4.2.2 Cross-Validation Sample

The summary of cross-validation is shown in Tables 6 through 8. Figures 16 through 18 also display the difference of Spearman rank-order correlation and the difference of PRESS for each of the subsamples by using different selection methods. The comparison between these two subsamples (labeled as sample 1 and sample 2) is presented in these figures.

As expected, the value of the rank-order coefficient obtained in the subsample of cross-validation is less than that attained in the total applicant sample. It is important to note that the value of difference in the cross-validation is much greater than the value of difference obtained in the applicant sample.

Furthermore, examining the values of PRESS across sample 1 and sample 2, it shows that the values in the hierarchical models are smaller than those in the single-factor models. These values of PRESS indicate that the prediction of using general and specific factors as predictors results in smaller amount of prediction error. In summary, comparisons across the subsamples, the results of rank correlation and PRESS are stable.

4.3 Effects of Selection Methods

In summary, Tables 3 through 5 and Figures 7 through 15 show that the prediction approach of using general and specific factor scores performs better than the method of using an estimated general factor score alone. By introducing the specific factor as predictor, this approach not only increases the precision of prediction in relation to the Spearman rank-order correlation, but also tends to produce efficient classification in terms of sensitivity, specificity, and proportion of correct decisions.

The results for cross-validation are similar to the results described above except for a few cases which show a negative difference. It is observed that the rank coefficients obtained from cross-validation are smaller than those obtained from the applicant sample but the increment is more notable for the results of cross-validation.

4.4 Effects of Selection Ratio and Sample Size

Given a particular sample size for the selected subjects, when the selection ratio increases, the proportion of successful individuals who are accepted (sensitivity) increases, the proportion of failing individuals who are rejected (specificity) decreases, and the proportion of correct decisions decreases. On the basis of this simulation, when the selection ratio decreases the difference in sensitivity between the two methods increases.

Comparing two sets of plots with the same number of selectees (N=250 or N=500) in Figure 10, the results show that the lower selection ratio leads to the larger increase of difference in sensitivity and to the smaller increase of difference in specificity

and proportion of correct decisions. The same phenomena as described above are observed in Figures 11 and 12. Figures 7 through 9 also present information as to how the selection ratio affects the results of classification.

Two different sample sizes for the selected subjects were studied to investigate the effect of sample size on prediction. The results show that the sample size for the selectees does not affect sensitivity, specificity, and proportion of correct decisions. In comparing two sets of plots with the same selection ratio in Figures 7 through 12, we observe that increasing the number of selectees does not change the results of classifications. The patterns of these results are very similar.

In the cross-validation, the sample size is decreased to 125 and 250 for each subsample to compute rank coefficient. As expected, the coefficients are smaller than those obtained from analyzing all applicants; in contrast, the difference in relation to the increment from the specific factor becomes more notable. It is concluded that selection ratio and sample size for the selected subjects have effects on rank coefficients. As sample size for the selected subjected or selection ratio increases, Spearman rank correlation coefficient increases. While the increment of difference depends on the increase in sample size for the selected subjects and the decrease of selection ratio.

4.5 Effects of Models

Three elements, R square, standardized regression slope, and determinacy, generate 12 ($2 \times 3 \times 2$) different conditions for each of the three cases. The following section addresses the effect for each factor separately.

4.5.1 R Square

It is important to note that high R^2 value will enhance the precision of selection and classification. The results show that the classifications from models with an R^2 value of 0.6 are more accurate than those with an R^2 value of 0.4. The high value of R^2 increases the value of the Spearman rank-order correlation coefficient in both the applicant sample and the cross-validation sample.

4.5.2 Standardized Regression Slopes

As expected, controlling for the factor loadings and R^2, the accuracy of classification by using multiple predictors is influenced by the value of the regression slope on the specific factor relative to that of the general factor. Considering models 1, 3, and 5 (see Figure 7), the values of sensitivity, specificity, and proportion of correct decisions are in descendant order as $\beta_g^* > \beta_s^*$ (model 1), $\beta_g^* \approx \beta_s^*$ (model 3), and $\beta_g^* < \beta_s^*$ (model 5). Using models 8, 10, and 12 as another example, the same pattern is observed in that model 8 has the highest values of sensitivity, specificity, and proportion of correct decisions.

When comparing the classification using only the general factor score to the classification using both general and specific factor scores, it may be noted that the greatest gain from the method of using general and specific factor scores is found in model 12, which contains a high value of β^*_s. If the values of R^2 and factor loadings are held constant, the effects of β^* could be examined among models 1, 3, and 5; or models 2, 4, and 6; or models 7, 9, and 11; or models 8, 10, and 12. It is observed that in the case of $\beta^*_g < \beta^*_s$, the prediction of using specific factors in addition to the general factor does much better than the prediction without using specific factors as predictors. The results of classification are similar to the results of Spearman rank-order correlation among these model comparisons.

The same phenomenon is observed in the results of cross-validation. Specifically, for models with $\beta^*_g < \beta^*_s$, the differences of Spearman rank-order correlations between $\hat{g} + \hat{s}$ and $\hat{g}_I$ are the greatest. Comparisons of $\hat{g} + \hat{s}$ to $\hat{g}_I$ in the case of g + 3 s and N=500, R=0.1 in Table 8, model 12 ($\beta^*_g < \beta^*_s$) results in an increase of 0.274 and 0.282 for Spearman rank-order correlation in sample 1 and sample 2, respectively; the increments for model 8 ($\beta^*_g > \beta^*_s$) are 0.068 and 0.059 in sample 1 and sample 2, respectively; the increments for model 10 ($\beta^*_g \approx \beta^*_s$) are 0.220 and 0.223 in sample 1 and sample 2, respectively.

4.5.3 Factor Loadings

When the regression slopes and the value of R^2 are held constant, the value of factor loadings (factor determinacy) will affect the prediction in terms of sensitivity, specificity, proportion of correct decisions, and Spearman rank-order correlation coefficient. A higher value of factor determinacy will improve the accuracy of prediction. The same effects of factor loadings can be observed in the results of cross-validation.

4 .6 The Increment of Prediction From the Specific Factors

The increment of prediction from the specific factors is evaluated by the classification and Spearman rank-order correlation coefficient for the applicants. The approach of cross-validation is also used for detecting the same effects on Spearman rank-order correlation coefficient.

4.6.1 The Gains in Classification

Looking at the results for $g + 1\ s$, it is noted that the range of difference for sensitivity is from 0.39% to 11.58%; the range of difference for specificity is from 0.09% to 6.48%; and the range of difference for proportion of correct decisions is from 0.16% to 6.48%. In the situations with a 10% selection ratio, the largest increment of sensitivity is about 11% in models 11 and 12, and the second largest increase is about 7% in models 5 and 6. There is a similarity among these four models. That is, the regression slope for the specific factor is greater than the weight for the general factor. It is observed that models 1, 2, 7, and 8 have a smaller amount of increase for sensitivity. When the specific

factor with low regression slope, the difference of sensitivity between the $\hat{g}_I$ method and the $\hat{g}$ + 1 $\hat{s}$ method decreases. For models with equivalent slopes for general and specific factors (models 3, 4, 9, and 10), the gain of sensitivity is about 3% to 4% for an R^2 value of 0.4 and 5% to 6% for an R^2 value of 0.6.

In Figure 7, each plot is presented by ordering the value of sensitivity, specificity, and proportion of correct decisions for $\hat{g}$ + 1 $\hat{s}$ among the twelve models. As expected, model 5 has the lowest value of sensitivity, specificity, and proportion of correct decisions. In contrast, model 8 has the highest value of sensitivity, specificity, and proportion of correct decisions. The prediction model with low R^2, low determinacy on specific factor, and low regression weight on specific factor tends to produce the least accurate results of selection and classification. If the specific factor plays an important role in the future prediction, the strategy of including the specific factor as predictor will improve the accuracy of selection and classification.

Similarly, the pattern of results for g + 2 s and g + 3 s is obtained as that for g + 1 s. Here, model 5 for g + 1 s (see Table 3) is chosen to be an example for demonstration. The difference shows an increase of 4% of the successful applicants who should be selected in the cases of 50% selection ratio, while in the cases using a 10% selection ratio, the difference increases 7% of the successful applicants who should be selected.

4.6.2 The Gains in Spearman Rank-Order Correlation Coefficient

The range of difference for the Spearman rank-order correlation is from 0.010 to 0.164 for the g + 1 s case, from 0.005 to 0.072 for the g + 2 s case, and from 0.002 to

0.132 for the $g + 3\ s$ case in the applicant sample. In addition, for the subsamples in cross-validation, the range of difference for the Spearman rank-order correlation is from 0.013 to 0.314 for the $g + 1\ s$ case, from –0.003 to 0.185 for the $g + 2\ s$ case, and from –0.008 to 0.282 for the $g + 3\ s$ case.

CHAPTER 5

REAL DATA APPLICATION

The findings reported from the simulation study point to benefits potentially obtainable from an effective selection and classification approach. The aim of this real data analysis is to provide an application of the use of factor scores as predictors in the context of selection and predictive validity. Analyses are performed for each of the nine Army Project A jobs (Campbell & Zook, 1991, 1992) and two jobs from the Marine Corps Job Performance Measurement (JPM[1]) Project using hands-on job performance as the criterion. The results of classification are evaluated by Spearman rank-order correlation coefficient, sensitivity, specificity, and proportion of correct decisions.

5.1 The Army Project A and the Marine Corps JPM Data

The Armed Services Vocational Aptitude Battery (ASVAB) is the test battery which the United States Military Services have used since 1976 to determine the cognitive qualification of applicants for service. The battery serves both to determine whether applicants meet minimum enlistment standards and to aid in determining the specialty area in which an applicant might most benefit from advanced training. In the Army Project A and the Marine Corps JPM, additional tests have been used to extend the range of abilities covered by the ASVAB.

[1]The Marine Corps JPM data have been kindly made available by Neil B. Carey at the U. S. Marine Corps.

5.1.1 Sample

The data analyzed for this report are from the nine different jobs of Batch A of the Army Project A Concurrent Validity Study and two jobs of the Marine Corps JPM data. Table 9 lists the nine Army jobs (Infantryman, Cannon Crewman, Tank Crewman, Radio Operator, Vehicle Mechanic, Motor Transport, Administrative, Medical, and Military Police) and two Marine jobs (Helicopter Mechanic and Automotive Mechanic) and gives the number of soldiers included in the present analyses.

5.1.2 Variables

A listing of predictor measures is given in Table 10. It lists 10 ASVAB subtests, 12 Army Project A subtests, and 8 subtests from the Enhanced Computer Administration Test (ECAT).

The Armed Services Vocational Aptitude Battery (ASVAB) includes ten tests: General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Numerical Comprehension, Coding Speed, Auto and Shop Information, Math Knowledge, Mechanical Comprehension, and Electronics Information. This battery measures reasoning, spatial visualization, psychomotor abilities, and working memory.

The twelve Army Project A subtests include six paper-and-pencil spatial tests: Assembling Objects, Maps, Mazes, Object Rotation, Orientation, and Figural Reasoning; and six computerized perceptual/psychomotor tests: Target Tracking Test 1, Target Tracking Test 2, Target Identification Test-Time, Target Identification Test-Hits, Memory Search Test-Time, and Memory Search Test-Hits.

The Enhanced Computer Administration Test (ECAT) battery consists of nine tests. Three of them are cognitive ability tests that require computer administration: Integrating Details, Mental Counters, and Sequential Memory. Three tests are psychomotor tests reproduced from the Army Project A: One-hand Tracking, Two-hand Tracking, and Target Identification. Three of the tests, Assembling Objects, Spatial Orientation, and Figural Reasoning, are computer-administered versions of Project A paper-and-pencil spatial tests (Wolfe, 1994).

ASVAB and the extended Army Project A subtests scores are available for the Army Project A while ECAT and ASVAB scores are available for the Marine Corps JPM data. ASVAB scores are available for the entire applicant population. The hands-on job performance criterion scores are available only for the selected population and differs across jobs.

5.2 Structural Models

In the Army Project A and the Marine Corps JPM data, additional tests have been similarly used to extend the range of abilities covered by the ASVAB. In both cases, the tests have a structure similar to the hierarchical model described earlier. According to the modeling of Gustafsson and Muthén (1994), the most important factors are Gf (general factor interpreted as fluid intelligence), Gc (crystallized intelligence), Gv/Mech (visual perception and mechanical knowledge), Speed (perceptual speed factor), Math (math knowledge), and Psymotor (general psychomotor speed).

The current analyses use the model from the applicant population factor solution of Gustafsson and Muthén (1994), which was based on the ten ASVAB subtests and twelve extra variables from Project A. This solution was obtained from an analysis of a covariance matrix obtained by a standard Pearson-Lawley adjustment made to the covariance matrix for the sample of all nine Army Project A jobs (Gustafsson & Muthén, 1994; Muthén & Gustafsson, 1995). The same model is also applied to the Marine Corps JPM data which includes the ten ASVAB subtests and eight ECAT variables.

Two models, the single-factor model and the Gustafsson-Muthén hierarchical latent variable model, are used as the differential selection method for this study. For the single-factor model, there is only a general factor (g_1) which influences the performance of examinees on all tests. The hierarchical model of abilities contains both general and specific factors. The factor structure for ASVAB and Project A subtests was found to be very close to that of the factor structure for ASVAB and ECAT subtests (Gustafsson & Muthén, 1994). Tables A1 and A2 (in Appendix A) show the factor structures for the Army Project A and the Marine Corps JPM datasets, respectively.

The hands-on job performance variable is regressed on the six factors defined in the Gustafsson-Muthén model (1994). Based on the missing data theory and the Pearson-Lawley adjustment, the regression slopes of the criterion on the factors are estimated by the regression method using the job samples. The factor scores are calculated for each individual as the predictors of job performance. Because the factor score coefficients are estimated from the applicant population factor model, the selective nature of the job samples is also considered.

For each job sample, both the Gustafsson-Muthén model and the single-factor model are obtained so we can compare the predictive validities across jobs. Furthermore, the predicted criterion scores can be estimated by these two different approaches. This report investigates whether the specific factors make a greater difference in predicting hands-on performance than the general factor does for each job.

5.3 Procedure of Analyses

Based on the strategy of the artificial data analyses and the available data patterns, the following analysis procedure is carried out:

Step 1. Obtain the measurement models based on the Gustafsson-Muthén model and the single-factor model;

Step 2. Estimate factor scores ($\hat{g}+5\hat{s}$ vs. $\hat{g}_1$) using the regression method (Lawley & Maxwell, 1971);

Step 3. Estimate the factor score determinacies;

Step 4. Specify the cutoff score (the top 50 %) based on $\hat{g}_1$ for obtaining a selected sample N_S;

Step 5. Estimate the regression slopes (β_g and β_S) for general and specific factors with the hands-on job performance (y) using the selected sample (N_S);

Step 6. Estimate the predicted hands-on job performance ($\hat{y}$) using the estimated factor scores and the estimated regression coefficients;

Step 7. Obtain a selected sample based on $\hat{y}$;

Step 8. Classify the true successful subjects based on the observed hands-on job performance (y);

Step 9. Compare rank correlation between the observed hands-on job performance (y) and the predicted hands-on job performance ($\hat{y}$);

Step 10. Perform sensitivity and specificity analyses to compare the results of different prediction methods.

In order to obtain the method corresponding to the simulation analyses, the top 50% of each job sample based on the observed hands-on job performance are viewed as true successful candidates in Figure 2 terms. Because of this, the analyses of sensitivity, specificity, and proportion of correct decisions can be carried out.

As will be seen, specific factors are not all significant for all 11 of the jobs. An alternative way of estimating the predicted criterion scores is to use the Gf factor in addition to the significant specific factors as predictors. The notation of $\hat{g} + 5\hat{s}$ means that the predictors contain the Gf factor and the 5 specific factors while the notation of $\hat{g} + \hat{s}^*$ means that the predictors include the Gf factor and the specific factors which are significant. Because the regression equations differ across the 11 jobs, the approach of $\hat{g} + \hat{s}^*$ has a diverse combination of predictors for each job. The notation of $\hat{g}_I$ denotes that the general factor defined in the single-factor model is the only predictor. Thus, there are three selection methods for classifying individuals.

5.4 Results

Descriptive statistics for the estimated factor scores and the hands-on job performance scores are shown in Table 11. Table 12 shows the estimates of the standardized prediction equation and the adjusted R square for each job sample. Table 13 displays the estimates of the standardized prediction equation for the top 50% of each job sample. The results of the Spearman rank-order correlation are shown in Table 14. The classifications in different combinations of selection methods and job samples are summarized in Table 15.

5.4.1 Factor Determinacy

Determinacies for the hierarchical factor model are shown in Table 11. In the Army Project A data, the determinacies for Gf, Gc, Gv/Mech, Speed, Math, and Psymotor are 0.927, 0.900, 0.865, 0.840, 0.776, and 0.634, respectively. In the Marine Corps JPM Automotive Mechanic data, the determinacies for Gf, Gc, Gv/Mech, Speed, Math, and Psymotor are 0.935, 0.887, 0.858, 0.853. 0.747, and 0.883, respectively. In the Marine Corps JPM Helicopter Mechanic data, the determinacies for Gf, Gc, Gv/Mech, Speed, Math, and Psymotor are 0.946, 0.902, 0.861, 0.859, 0.742, and 0.916, respectively.

In summary, the determinacy for Gf is quite robust but the determinacy for Math drops below 0.8 among the Army Project A and the Marine Corps JPM. It is important to note that the determinacy for Psymotor is only 0.634 in the Army Project A while the determinacy for Psymotor is much higher in the Marine Corps JPM. The results show that the ECAT subtests tend to give a more accurate measure of Psymotor than the Army

Project A subtests, which do not capture Psymotor very well in terms of determinacy. In general, the instruments for measuring math knowledge do not perform as well as those for measuring other abilities.

5.4.2 Standardized Regression Equation and the Adjusted R Square

The job sample (*N*) and the top 50% of the job sample (N_S) are used for estimating the regression equations. Table 12 shows the standardized estimates for the prediction equation with the total job sample. In this table, the adjusted R^2 column represents the adjusted value using the six factors as predictors across the 11 jobs. It is convenient to compare the predictive strength and the regression slopes across jobs with respect to the adjusted R^2 value and the standardized regression equation. The hierarchical model shows that the Gf factor has a strong influence on criterion for all 11 jobs. The Gc factor is important for Radio Operator, Medical, and Automotive Mechanic. The Gv/Mech factor has a significant effect on hands-on job performance for all jobs except Administrative. The Speed factor is only important for Radio Operator and Vehicle Mechanic. The Math factor and the Psymotor factor are significant for Military Police and Vehicle Mechanic, respectively. The results of the adjusted R^2 show that Cannon Crewman has the lowest value. The highest value of R^2 is about 0.57 for Automotive Mechanic. It is observed that the same structure of cognitive ability gives different strengths of prediction in different job performances. The Marine jobs show the highest predictive strength.

The estimates of the standardized prediction equation for the top 50% of each job sample are shown in Table 13. The selected sample was chosen from the top 50% of

ranking by $\hat{g}_I$. The results show that the Gf factor is not important for Cannon Crewman, Tank Crewman, Radio Operator, and Vehicle Mechanic in the selected sample. The Gc factor has a significant effect on Radio Operator, Motor Transport, and Automotive Mechanic. It is of interest to note that Gc is important for the top 50% of Motor Transport sample, but Gc is not an important predictor for the Motor Transport sample. The Speed factor is important for Vehicle Mechanic for the highest-ranked selectees. In these selected samples, the Gv/Mech factor becomes more important than the Gf factor for Cannon Crewman, Tank Crewman, Vehicle Mechanic, Motor Transport, and Automotive Mechanic. The significant effect on Gv/Mech is found in eight jobs. Math factor is significant for Military Police, but Psymotor factor has no significant effect on any of the jobs. None of the six factors are important for predicting the top 50% of Cannon Crewman, so the alternative method for $\hat{g}+\hat{s}^*$ is to use the Gf factor only. These estimated regression coefficients from the selected job samples are used for the following analyses (Spearman rank-order correlation and classification).

5.4.3 The Spearman Rank-Order Correlation Coefficient

The results of the Spearman rank-order correlations between the predicted criterion scores and the observed hands-on job performance are estimated by using the three methods defined above (see section 5.3). A higher value of the coefficient means the rank-ordered $\hat{y}$ (predicted criterion) based on a certain method is closer to the ranking by the observed hands-on job performance. As shown in Table 13, the method of using $\hat{g}+5\hat{s}$ performs better than the methods of using $\hat{g}+\hat{s}^*$ and $\hat{g}_I$. Most of the rank

correlations between the observed criterion and the predicted criterion by $\hat{g}+\hat{s}^*$ are higher than those using only the general factor ($\hat{g}_1$) except for the jobs of Cannon Crewman, Radio Operator, and Medical. Two sets of the coefficients for $\hat{g}+\hat{s}^*$ and $\hat{g}_1$ are nearly identical for each of these three jobs. Therefore, the general ability is an important indicator for predicting the hands-on performance in Cannon Crewman, Radio Operator, and Medical jobs.

It is of interest to note that there is no significant factor found in the Cannon Crewman prediction (see Table 13). Thus, the comparison of $\hat{g}+\hat{s}^*$ and $\hat{g}_1$ can be seen as the comparison of the Gf factor and the general factor for Cannon Crewman. The results indicate that for this jobs, the Gf factor is similar to the general factor in predicting the hands-on performance. The results of Cannon Crewman and Medical demonstrate that in practice, the situations with varied prediction strength (R^2) are still in favor of the general ability as the predictor, so the difference among the three rank coefficients is tiny.

The results for the two Marine jobs show another pattern: the specific factors enhance the precision of ranking individuals dramatically. This is to be expected given that the specific factors strengthen the prediction. We also note that the value of R^2 is also high in Marine jobs. As the results of Monte Carlo study indicate that the high value of R^2 tends to have high value of the Spearman rank-order correlation.

5.4.4 Sensitivity , Specificity, and Proportion of Correct Decisions

Table 15 displays the job classifications with respect to the sensitivity, specificity, and the proportion of correct decisions. When the value of R^2 is taken into account, it is

observed that the cases with higher values of R^2 perform consistently with the results from the simulation analyses. For example, in Automotive Mechanic with an R^2 value of 0.57, $\hat{g}+5\hat{s}$ performs better than $\hat{g}+\hat{s}^*$ and $\hat{g}_1$ in terms of the values of sensitivity, specificity, and the proportion of correct decisions. The difference value is about 4.32%.

It is important to note that the regression coefficients shown in Table 13 are from the top 50% of the job sample. For Cannon Crewman, the selection method of $\hat{g}$ represents the Gf factor without any specific factors because there is not any significant factor found in this regression. The classification of this job shows that general ability is the best predictor. It is also observed that including the predictors of the specific factors hamper the precision of classifications. The extreme results for Cannon Crewman show that $\hat{g}+5\hat{s}$ performs worst. Because of the lack of significant predictors and the low value of R square, the results for Cannon Crewman are regarded with suspicion.

The comparison between $\hat{g}+5\hat{s}$ and $\hat{g}_1$ shows that the general factor gives better predictive classification in Infantryman, Cannon Crewman, Tank Crewman, and Radio Operator. Results from the rest of the jobs shed light on the use of specific factors. It is expected that adding the significant factors to Gf will lead to a similar result obtained from the six factors, and perform better than the general ability does. No certain pattern is found in this real data application, so this assumption has not been validated. However, it is observed that the results of sensitivity, specificity, and proportion of correct decisions are almost identical for $\hat{g}+5\hat{s}$ and $\hat{g}+\hat{s}^*$ in Helicopter Mechanic and Automotive Mechanic jobs.

Although the R^2 value for Vehicle Mechanic is only 0.16, the gain of sensitivity from $\hat{g} + 5\hat{s}$ is about 3%. In practice, the increase of 3% is a remarkable improvement for the selection. Assume a selection setting with 10,000 applicants and 50% selection ratio. Here, an increase of 3% for sensitivity implies that about 150 additional successful subjects out of the 5,000 successful applicants would be accepted by the method of using specific factors in addition to the general factor. The other findings show that the classification of Tank Crewman and Radio Operator is best predicted by the general ability. These results imply that the specific skills is not required for Tank Crewman and Radio Operator.

5.4.5 Profile Description

A description of the profile for the top 100 individuals in the observed hands-on performance for each job is given in the appendix B. The variable TOP represents the rank-ordered hands-on performance; when the subject is in the top 100 rank-ordered category then TOP will be coded to 1, otherwise TOP will be assigned to 0. NEW, SIG, and GEN variables represent that the methods of $\hat{g} + 5\hat{s}$, $\hat{g} + \hat{s}^*$, and $\hat{g}_1$ are used for predicting hands-on performance, respectively. These three variables are sorted from highest to lowest and given rank-ordered values for each subject; then, each variable is classified into different groups (e.g., group 1 includes the rank-order values from 1 to 100; group 2 ranging from 101 to 200; group 3 ranging from 301 to 400; group 4 ranging from 401 to 500; group 5 ranging from 501 to 600; group 6 ranging from 601 to 700). A

two-way table shows both the frequency and percentage crosstabs for two ordered variables (TOP by NEW, TOP by SIG, or TOP by GEN) for each job sample.

Looking at the result for job 6 (Motor Transport), for example, the highest rank-order individuals (TOP = 1) are classified by three approaches: NEW ($\hat{g}+5\hat{s}$), SIG ($\hat{g}+\hat{s}^{*}$), and GEN ($\hat{g}_{1}$), separately. For the $\hat{g}+5\hat{s}$ method, 43 subjects are classified into group 1, 23 subjects are ranked in group 2, 16 subjects are in group 3, 12 subjects are in group 4, and 6 subjects are in group 5. For the $\hat{g}+\hat{s}^{*}$ approach, 41 individuals are classified into group 1. Based on the classification of $\hat{g}_{1}$, there are 35 subjects in group 1. If there are 100 vacancies in Motor Transport job sample, $\hat{g}+5\hat{s}$ and $\hat{g}+\hat{s}^{*}$ will result in more accurate decisions than $\hat{g}_{1}$. Similarly, 57 subjects having low performance are selected by $\hat{g}+5\hat{s}$, 59 low-performance subjects are predicted to be successful by $\hat{g}+\hat{s}^{*}$, and 65 subjects are misclassified as potential candidates by $\hat{g}_{1}$. It is observed that the selection method of using general and specific factors tends to accept more subjects who will be successful and tends to avoid the error of hiring subjects who will fail in future performance.

CHAPTER 6

DISCUSSION

6.1 Discussion of the Monte Carlo Study

Results from the simulation study suggest that using specific factors in addition to a general factor as predictors gives better selection decisions. Because the true criterion is available for each applicant, the classification of applicants based on the true criterion and cutoff (selection ratio) is used for evaluating the success rate of the classifications on these two selection methods.

It should be pointed out that these findings, as well as the findings of Hsu (1995), support the notion that sample size for selected subjects does not affect the prediction. The results of sensitivity indicate that the specific factors play an important role in the context of prediction and selection. It should also be noted that the percentage of sensitivity increase rate multiplied by the number of successful subjects gives information about the number of acquired successful subjects by the new selection approach. Figure 19 shows how a 4% increase in the rate of sensitivity affects the decision of accepting successful applicants. When the sample size of applicants is 100,000 with a 50% selection ratio, 4% increase in the sensitivity will lead to an increase of accepting 2,000 successful subjects. There are about 650,000 applicants taking the ASVAB for military enlistment every year. This means that a 4% of the increase in sensitivity with a 50% selection ratio will increase the correct selection of 13,000 successful subjects. The results of sensitivity also show that the most selective situation in terms of low selection

ratio (10%) results in about twice the increment in the sensitivity as compared to the high 50% selection ratio. This implies that the method of using specific factors has potential contributions, especially in the setting with a large applicant sample.

This report has examined the effects of R^2, regression slopes, and factor determinacies in predictive accuracy in regard to identification of applicants' future performance. It is clear that the hierarchical model in the condition of high R^2, with more emphasis on the general factor and high determinacies for specific factors, results in more accurate identification of applicants than the other contrast models when the determinacy for the general factor is held constant. Another important finding is that the model with high R^2, high regression slope on specific factors, and high determinacies for specific factors leads to the greatest increase in value on sensitivity when compared to the single factor model. This finding is verified in the real data application which presents evidence for classification of Helicopter Mechanic and Automotive Mechanic jobs.

6.2 Discussion of the Real Data Application

There are some restrictions in the analyses of the real data application. First, the subtest scores are limited to part of the selected sample so the classification is not carried out in the non-selected sample, and the evaluating methods are carried out for the selected sample. Second, cross-validation is not carried out in real data analyses because of sample size.

The results show that for the Army Project A data, the Gf, Gc, Gv/Mech, and Speed factor determinacies are quite high, which implies that the factor scores are

reliable. In the Marine Corps JPM data, the Math factor determinacy is about 0.75 but the rest of the factor scores are quite reliable. Based on the values of determinacies, we can conclude that the ECAT provides better measure for Psymotor; in contrast, Math ability is not well defined by the ASVAB, ECAT, and Army Project A subtests. These results suggest that more investigation and development of subtests for measuring Math ability are needed.

Among Vehicle Mechanic, Motor Transport, Administrative, Helicopter Mechanic, and Automotive Mechanic jobs, the specific factors do increase the rank coefficient. The increment of Spearman rank order correlation coefficient implies that these jobs may require more specific abilities. For the prediction of Cannon Crewman, the general factor gives slightly better performance than does the Gf factor in addition to the Gv/Mech factor. In the case of Cannon Crewman, the results show that the six factors only explain 5% of the variance of the job performance. The regression gives a bad fit, so any conclusion about this job is limited. It is most likely that adding noncognitive predictors to the ability would improve the prediction of hands-on performance for Cannon Crewman.

For Radio Operator, the regression equation shows that the Gf factor, the Gc factor, the Gv/Mech factor, and the Speed factor are important predictors. For the Medical job, the regression equation shows that three factors, the Gf factor, the Gc factor, and the Gv/Mech factor, are the most important predictors. The Spearman rank-order coefficients, however, show that there is not much difference among the three sets of

values in Radio Operator and Medical jobs. It is observed that the additional specific factors do not enhance the predictive validity for these two jobs.

Also, fluid and crystallized ability, as measured by intelligence or aptitude tests, play an important part in academic and occupational success. This fact demonstrates that these abilities are general and are not specific to the tests themselves. The hands-on performance of the Administrative job is better predicted by the fluid ability than by the general ability. This points to the advantage of using the hierarchical model for differentiating the structure of ability. The prediction equation of Vehicle Mechanic indicates that the Gv/Mech factor is more important than the Gf factor. This fact validates that the scenario $\beta_g^* < \beta^*_s$, created in the simulation, is close to the practical situation.

For the Marine jobs, the Gv/Mech factor for Helicopter Mechanic and Automotive Mechanic and the Gc factor for Automotive Mechanic significantly improve the Spearman rank-order correlations. The differences of the rank correlations between $\hat{g} + 5\hat{s}$ (or $\hat{g} + \hat{s}^*$) and $\hat{g}_1$ are about 0.1. The comparison between these two approaches with a 50% selection ratio shows that there is an increase of about 4% for sensitivity, specificity, and proportion of correct decisions. It is of interest to note that for Automotive Mechanic the standardized regression slope for the Gv/Mech factor (0.355) is close to that for the Gf factor (0.343), while the standardized regression slope for the Gc factor (0.135) is smaller than that for the Gf factor. For Helicopter Mechanic, the standardized regression equation shows that the regression slope for the Gf factor (0.404)

is much higher than those for the specific factors (e.g., 0.295 for Gv/Mech). These results provide evidence that specific factors have valuable contributions in selection and prediction.

In summary, most results of the real data analyses agree with the results of the simulation study. The results suggest that the Army and the Marine Corps can improve the prediction of job performance by adding specific factor scores as predictors.

6.3 Limitations, Implications, and Recommendations

The results of the real data application point to the limitations of the Monte Carlo study. The values of R square are set at 0.4 and 0.6 in the Monte Carlo study. In practice, the values of R square for Infantryman, Tank Crewman, and Vehicle Mechanic are below 0.2. For Cannon Crewman the R square is only 0.05. It would be useful to study models with much lower R square value than in the Monte Carlo study to examine the effects of selection methods. Furthermore, Short (1990) mentioned that the portion of items influenced by specific factors was found to be more influential in obtaining reliable factor scores than the number of items overall, so the number of subtests for each factor should be considered with respect to the issue of factor determinacy. The number of subtests for each specific factor is designed to be constant in the Monte Carlo study. Since unequal number of subtests for each specific factor could affect the reliability of factor score, it could be included in future studies.

It is important to recognize that the scientific and practical utility of criterion validation depends as much on the measurement of the criterion as it does on the quality

of the measuring instrument itself. Thus, in many different types of training programs, much effort and expense goes into the development of a test for predicting who will benefit from the program in terms of subsequent job performance. In the area of special education or prevention programs, the use of factor scores will lead to more accurate identification and classification for students regarding future performance. Because of the limited resources, how to determine which student receives treatment and which student is ineligible for treatment represents an important issue. For example, the traditional approaches to the identification of reading problems require comparison of standardized achievement and IQ measures. Factor scores can be applied to differentiating skill areas (e.g., general reading ability, spelling skill, and vocabulary skill). When evaluating these differences in the light of the theory of reading problems, the profiles of factor scores will lead to the prediction that individuals with less spelling skill tend to be the risk group for poor academic outcome. The use of specific factors can be applied to the specialized training program. For instance, the program for musical talents will admit subjects with musical aptitude (e.g., special talent in pitch, rhythm, or tone) rather than those with high academic achievement or IQ score.

Predictive validity has been used in the fields of psychology and education mainly for analyzing the validity of certain types of tests and selection procedures. Many standardized intelligence tests, achievement tests, and ability tests were designed to provide information for the selection and placement of students and for comparing students, schools, and school districts with one another. Since many children have been misclassified by standardized tests and on this basis have been assigned to programs with

minimal content, low expectations, and restricted (rather than enriched) teaching approaches, the link between assessment and intervention becomes more important than the issues of placement and prediction. A new type of test should provide diagnostic information about students' preconceptions, learning strategies, and metacognitive and affective thought processes. The approach of using factor scores should contribute to effective diagnostic analysis.

The most difficult problem in social science studies is that the hypothetical concepts and constructs are not directly measurable. Although such concepts and constructs, or latent variables, cannot be directly measured, a number of variables can be used to measure various aspects of these latent variables more or less accurately. We may regard the observed variables as indicators of the latent variables. Each indicator has a relationship with the latent variable, but if we take one indicator alone to measure the latent variable, we would obtain a biased measurement. Using several indicators of each latent variable gives a better measurement of the latent variable. Another reason for using latent variables in behavioral and socioeconomic studies is that most of the measurements employed contain sizable errors of measurement, which, if not taken into account, can cause severe bias in the results. Errors of measurement arise because of imperfection in the various measurement instruments that are used to measure people's behavior, attitudes, feelings, and motivations. Even if we could construct valid measurement for these traits, it is usually impossible to obtain perfectly reliable variables. In practice, using factor scores instead of raw score or composite scores will be proper for profile analysis.

As an alternative to dollar value units, utility gains can be expressed as percentage increases in output. It is easy to realize that such output increases imply large dollar values, and there is a tendency to regard dollar estimates as indicating greater utility than the percentage-increase estimates. It should be emphasized that the use of specific factor scores deserves special recognition as an improvement in selection and prediction. Given the findings, these results suggest that the specific factor scores may facilitate the identification of students with reading disabilities and learning impairments in special populations.

In the Army Project A data and the Marine Corps JPM data, each job has different hands-on job performance so the use of factor score can obtain unbiased estimates of regression slopes for each job. These equations can then be applied to the new enlistment cohort for selection and classification reference. Moreover, based on the different prediction equations, the different predicted hands-on performance for each of the different jobs could be computed for each individual. The information of predicted hands-on performance could be used for matching people to jobs.

Previous research (Campbell, McHenry, & Wise, 1990) using the Army Project A data showed that some dimensions of job behavior, such as physical fitness and military bearing, are better predicted by noncognitive than cognitive predictors, and are better predicted by some noncognitive predictors than by others. Hogan (1991), Tett et al. (1991), and Schmit et al. (1995) suggested that some dimensions of job behavior can be predicted reliably by personality measures. Although the use of personality as a predictor in personnel selection has not been substantially successful in the past, Irving (1993) has

suggested that personality measures are related to performance criteria which are unrelated to cognitive ability when the traits measured are conceptually related to these criteria. It seems that personality measures may predict job performance dimensions which cannot be predicted by cognitive ability measures. The use of personality measures in personnel selection may be warranted when a careful job analysis is undertaken to determine which performance dimensions may be related to personality traits.

As a final comment, the results of this report should be viewed with caution. Only a limited number of situations are included in the Monte Carlo study. Further research using data on additional occupations is necessary to corroborate the findings presented here. Despite its limitations, this report has shed some light on the nature of the relationship between aptitude tests and job performance and may stimulate additional research on an important topic in the area of prediction. In summary, the results indicate that the method of utilizing specific factor information studied in this research is valuable for personnel selection.

Table 1
Design of the Simulation Study

Data

	Sample Size for the Selected Subjects	
Selection Ratio	250	500
0.1	Case × Factor	Case × Factor
0.5	Case × Factor	Case × Factor

Models

Case		$g + 1\ s$		$g + 2\ s$		$g + 3\ s$	
Factor		R^2		R^2		R^2	
β^*	λ_s	0.4	0.6	0.4	0.6	0.4	0.6
$\beta_g^* > \beta_s^*$	Low λ_s	Model 1	Model 7	Model 1	Model 7	Model 1	Model 7
	High λ_s	Model 2	Model 8	Model 2	Model 8	Model 2	Model 8
$\beta_g^* \approx \beta_s^*$	Low λ_s	Model 3	Model 9	Model 3	Model 9	Model 3	Model 9
	High λ_s	Model 4	Model 10	Model 4	Model 10	Model 4	Model 10
$\beta_g^* < \beta_s^*$	Low λ_s	Model 5	Model 11	Model 5	Model 11	Model 5	Model 11
	High λ_s	Model 6	Model 12	Model 6	Model 12	Model 6	Model 12

Note: Case represents $g + 1\ s$, $g + 2\ s$, and $g + 3\ s$.
Factor represents R^2, β^*, and λ_s.

Table 2
Model Parameter Values ($g + 1\ s$)

Parameter	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Factor Loading Λ^*	g	s_1	g	s_1	g	s_1	g	s_1	g	s_1	g	s_1
x_1	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00
x_2	0.75	0.00	0.75	0.00	0.75	0.00	0.75	0.00	0.75	0.00	0.75	0.00
x_3	0.70	0.00	0.70	0.00	0.70	0.00	0.70	0.00	0.70	0.00	0.70	0.00
x_4	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00
x_5	0.65	0.00	0.65	0.00	0.65	0.00	0.65	0.00	0.65	0.00	0.65	0.00
x_6	0.70	0.27	0.70	0.45	0.70	0.27	0.70	0.45	0.70	0.27	0.70	0.45
x_7	0.50	0.36	0.50	0.54	0.50	0.36	0.50	0.54	0.50	0.36	0.50	0.54
x_8	0.60	0.18	0.60	0.36	0.60	0.18	0.60	0.36	0.60	0.18	0.60	0.36
x_9	0.35	0.45	0.35	0.63	0.35	0.45	0.35	0.63	0.35	0.45	0.35	0.63
x_{10}	0.25	0.54	0.25	0.72	0.25	0.54	0.25	0.72	0.25	0.54	0.25	0.72
Regression Slope												
β	0.60	0.40	0.60	0.40	0.49	0.55	0.49	0.55	0.35	0.67	0.35	0.67
β^*	0.54	0.32	0.54	0.32	0.45	0.45	0.45	0.45	0.32	0.55	0.32	0.55
$var(\eta)$	1.00	0.80	1.00	0.80	1.00	0.81	1.00	0.81	1.00	0.81	1.00	0.81
Factor Determinacy	0.94	0.70	0.94	0.85	0.94	0.71	0.94	0.86	0.94	0.71	0.94	0.86
$var(y)$	1.22		1.22		1.21		1.21		1.21		1.21	

Parameter	Model 7		Model 8		Model 9		Model 10		Model 11		Model 12	
Factor Loading Λ^*	g	s_1	g	s_1	g	s_1	g	s_1	g	s_1	g	s_1
x_1	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00
x_2	0.75	0.00	0.75	0.00	0.75	0.00	0.75	0.00	0.75	0.00	0.75	0.00
x_3	0.70	0.00	0.70	0.00	0.70	0.00	0.70	0.00	0.70	0.00	0.70	0.00
x_4	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00	0.80	0.00
x_5	0.65	0.00	0.65	0.00	0.65	0.00	0.65	0.00	0.65	0.00	0.65	0.00
x_6	0.70	0.27	0.70	0.45	0.70	0.27	0.70	0.45	0.70	0.27	0.70	0.45
x_7	0.50	0.36	0.50	0.54	0.50	0.36	0.50	0.54	0.50	0.36	0.50	0.54
x_8	0.60	0.18	0.60	0.36	0.60	0.18	0.60	0.36	0.60	0.18	0.60	0.36
x_9	0.35	0.45	0.35	0.63	0.35	0.45	0.35	0.63	0.35	0.45	0.35	0.63
x_{10}	0.25	0.54	0.25	0.72	0.25	0.54	0.25	0.72	0.25	0.54	0.25	0.72
Regression Slope												
β	0.60	0.40	0.60	0.40	0.49	0.55	0.49	0.55	0.35	0.67	0.41	0.63
β^*	0.67	0.40	0.67	0.40	0.54	0.55	0.54	0.55	0.39	0.67	0.39	0.67
$var(\eta)$	1.00	0.80	1.00	0.80	1.00	0.81	1.00	0.81	1.00	0.81	1.00	0.81
Factor Determinacy	0.94	0.70	0.94	0.85	0.94	0.71	0.94	0.86	0.94	0.71	0.94	0.86
$var(y)$	0.81		0.81		0.81		0.81		0.81		0.81	

Table 2
Model Parameter Values ($g + 2\,s$)

Parameter	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
Factor Loading Λ^*	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2
x_1	0.80	0.27	0.00	0.80	0.45	0.00	0.80	0.27	0.00	0.80	0.45	0.00	0.80	0.27	0.00	0.80	0.45	0.00
x_2	0.75	0.36	0.00	0.75	0.54	0.00	0.75	0.36	0.00	0.75	0.54	0.00	0.75	0.36	0.00	0.75	0.54	0.00
x_3	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00
x_4	0.80	0.18	0.00	0.80	0.45	0.00	0.80	0.18	0.00	0.80	0.45	0.00	0.80	0.18	0.00	0.80	0.45	0.00
x_5	0.65	0.54	0.00	0.65	0.72	0.00	0.65	0.54	0.00	0.65	0.72	0.00	0.65	0.54	0.00	0.65	0.72	0.00
x_6	0.70	0.00	0.19	0.70	0.00	0.38	0.70	0.00	0.18	0.70	0.00	0.36	0.70	0.00	0.18	0.70	0.00	0.36
x_7	0.50	0.00	0.25	0.50	0.00	0.38	0.50	0.00	0.24	0.50	0.00	0.36	0.50	0.00	0.24	0.50	0.00	0.36
x_8	0.60	0.00	0.13	0.60	0.00	0.32	0.60	0.00	0.12	0.60	0.00	0.30	0.60	0.00	0.12	0.60	0.00	0.30
x_9	0.35	0.00	0.32	0.35	0.00	0.44	0.35	0.00	0.30	0.35	0.00	0.42	0.35	0.00	0.30	0.35	0.00	0.42
x_{10}	0.25	0.00	0.38	0.25	0.00	0.51	0.25	0.00	0.36	0.25	0.00	0.48	0.25	0.00	0.36	0.25	0.00	0.48
Regression Slope																		
β	0.40	0.30	0.30	0.40	0.30	0.30	0.29	0.32	0.49	0.29	-0.32	0.49	0.30	0.40	0.60	0.30	0.40	0.60
β^*	0.49	0.33	0.23	0.49	0.33	0.23	0.36	0.36	0.37	0.36	0.36	0.37	0.32	0.39	0.39	0.32	0.39	0.39
$var(\eta)$	1.00	0.80	0.40	1.00	0.80	0.40	1.00	0.81	0.36	1.00	0.81	0.360	1.00	0.81	0.36	1.00	0.81	0.36
Factor Determinacy	0.92	0.68	0.53	0.90	0.84	0.71	0.92	0.68	0.51	0.90	0.85	0.68	0.92	0.68	0.51	0.90	0.85	0.68
$var(y)$	0.67			0.67			0.64			0.64			0.87			0.87		

Parameter	Model 7			Model 8			Model 9			Model 10			Model 11			Model 12		
Factor Loading Λ^*	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2	g	s_1	s_2
x_1	0.80	0.27	0.00	0.80	0.45	0.00	0.80	0.27	0.00	0.80	0.45	0.00	0.80	0.27	0.00	0.80	0.45	0.00
x_2	0.75	0.36	0.00	0.75	0.54	0.00	0.75	0.36	0.00	0.75	0.54	0.00	0.75	0.36	0.00	0.75	0.54	0.00
x_3	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00	0.70	0.45	0.00
x_4	0.80	0.18	0.00	0.80	0.45	0.00	0.80	0.18	0.00	0.80	0.45	0.00	0.80	0.18	0.00	0.80	0.45	0.00
x_5	0.65	0.54	0.00	0.65	0.72	0.00	0.65	0.54	0.00	0.65	0.72	0.00	0.65	0.54	0.00	0.65	0.72	0.00
x_6	0.70	0.00	0.19	0.70	0.00	0.38	0.70	0.00	0.18	0.70	0.00	0.36	0.70	0.00	0.18	0.70	0.00	0.36
x_7	0.50	0.00	0.25	0.50	0.00	0.38	0.50	0.00	0.24	0.50	0.00	0.36	0.50	0.00	0.24	0.50	0.00	0.36
x_8	0.60	0.00	0.13	0.60	0.00	0.32	0.60	0.00	0.12	0.60	0.00	0.30	0.60	0.00	0.12	0.60	0.00	0.30
x_9	0.35	0.00	0.32	0.35	0.00	0.44	0.35	0.00	0.30	0.35	0.00	0.42	0.35	0.00	0.30	0.35	0.00	0.42
x_{10}	0.25	0.00	0.38	0.25	0.00	0.51	0.25	0.00	0.36	0.25	0.00	0.48	0.25	0.00	0.36	0.25	0.00	0.48
Regression Slope																		
β	0.40	0.30	0.30	0.40	0.30	0.30	0.29	0.32	0.49	0.29	0.32	0.49	0.30	0.40	0.60	0.30	0.40	0.60
β^*	0.60	0.40	0.28	0.60	0.40	0.28	0.44	0.44	0.45	0.44	0.44	0.45	0.39	0.47	-0.47	0.39	0.47	0.47
$var(\eta)$	1.00	0.80	0.40	1.00	0.80	0.40	1.00	0.81	0.36	1.00	0.81	0.360	1.00	0.81	0.36	1.00	0.81	0.36
Factor Determinacy	0.92	0.68	0.53	0.90	0.84	0.71	0.92	0.68	0.51	0.90	0.85	0.68	0.92	0.68	0.51	0.90	0.85	0.68
$var(y)$	0.45			0.45			0.43			0.43			0.58			0.58		

Table 2
Model Parameter Values $(g + 3\ s)$

Parameter	Model 1				Model 2				Model 3				Model 4				Model 5				Model 6			
Factor Loading Λ^*	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3
x_1	0.80	0.45	0.00	0.00	0.80	0.36	0.00	0.00	0.80	0.45	0.00	0.00	0.80	0.36	0.00	0.00	0.80	0.45	0.00	0.00	0.80	0.36	0.00	0.00
x_2	0.75	0.27	0.00	0.00	0.75	0.45	0.00	0.00	0.75	0.27	0.00	0.00	0.75	0.45	0.00	0.00	0.75	0.27	0.00	0.00	0.75	0.45	0.00	0.00
x_3	0.70	0.27	0.00	0.00	0.70	0.54	0.00	0.00	0.70	0.27	0.00	0.00	0.70	0.54	0.00	0.00	0.70	0.27	0.00	0.00	0.70	0.54	0.00	0.00
x_4	0.80	0.36	0.31	0.00	0.80	0.36	0.31	0.00	0.80	0.36	0.32	0.00	0.80	0.36	0.32	0.00	0.80	0.36	0.32	0.00	0.80	0.36	0.32	0.00
x_5	0.65	0.36	0.31	0.00	0.65	0.63	0.23	0.00	0.65	0.36	0.32	0.00	0.65	0.63	0.24	0.00	0.65	0.36	0.32	0.00	0.65	0.63	0.24	0.00
x_6	0.70	0.00	0.23	0.00	0.70	0.00	0.39	0.00	0.70	0.00	0.24	0.00	0.70	0.00	0.40	0.00	0.70	0.00	0.24	0.00	0.70	0.00	0.40	0.00
x_7	0.50	0.00	0.54	0.25	0.50	0.00	0.54	0.44	0.50	0.00	0.56	0.24	0.50	0.00	0.56	0.42	0.50	0.00	0.56	0.24	0.50	0.00	0.56	0.42
x_8	0.60	0.00	0.00	0.25	0.60	0.00	0.00	0.38	0.60	0.00	0.00	0.24	0.60	0.00	0.00	0.36	0.60	0.00	0.00	0.24	0.60	0.00	0.00	0.36
x_9	0.35	0.00	0.00	0.38	0.35	0.00	0.00	0.47	0.35	0.00	0.00	0.36	0.35	0.00	0.00	0.45	0.35	0.00	0.00	0.36	0.35	0.00	0.00	0.45
x_{10}	0.25	0.00	0.00	0.44	0.25	0.00	0.00	0.51	0.25	0.00	0.00	0.42	0.25	0.00	0.00	0.48	0.25	0.00	0.00	0.42	0.25	0.00	0.00	0.48
Regression Slope																								
β	0.50	0.50	0.30	0.30	0.50	0.50	0.30	0.30	0.38	0.42	0.47	0.63	0.38	0.42	0.47	0.63	0.30	0.40	0.60	0.60	0.30	0.40	0.60	0.60
β^*	0.43	0.38	0.20	0.16	0.43	0.38	0.20	0.16	0.32	0.32	0.31	0.32	0.32	0.32	0.31	0.32	0.25	0.30	0.40	0.30	0.25	0.30	0.40	0.30
$var(\eta)$	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40	1.00	0.81	0.64	0.36	1.00	0.81	0.64	0.36	1.00	0.81	0.64	0.36	1.00	0.81	0.64	0.36
Factor Determinacy	0.91	0.62	0.72	0.61	0.90	0.79	0.76	0.75	0.91	0.62	0.72	0.61	0.90	0.79	0.76	0.75	0.91	0.62	0.72	0.61	0.90	0.79	0.76	0.75
$var(y)$	1.35				1.35				1.44				1.44				1.44				1.44			

Parameter	Model 7				Model 8				Model 9				Model 10				Model 11				Model 12			
Factor Loading Λ^*	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3	g	s_1	s_2	s_3
x_1	0.80	0.45	0.00	0.00	0.80	0.36	0.00	0.00	0.80	0.45	0.00	0.00	0.80	0.36	0.00	0.00	0.80	0.45	0.00	0.00	0.80	0.36	0.00	0.00
x_2	0.75	0.27	0.00	0.00	0.75	0.45	0.00	0.00	0.75	0.27	0.00	0.00	0.75	0.45	0.00	0.00	0.75	0.27	0.00	0.00	0.75	0.45	0.00	0.00
x_3	0.70	0.27	0.00	0.00	0.70	0.54	0.00	0.00	0.70	0.27	0.00	0.00	0.70	0.54	0.00	0.00	0.70	0.27	0.00	0.00	0.70	0.54	0.00	0.00
x_4	0.80	0.36	0.31	0.00	0.80	0.36	0.31	0.00	0.80	0.36	0.32	0.00	0.80	0.36	0.32	0.00	0.80	0.36	0.32	0.00	0.80	0.36	0.32	0.00
x_5	0.65	0.36	0.31	0.00	0.65	0.63	0.23	0.00	0.65	0.36	0.32	0.00	0.65	0.63	0.24	0.00	0.65	0.36	0.32	0.00	0.65	0.63	0.24	0.00
x_6	0.70	0.00	0.23	0.00	0.70	0.00	0.39	0.00	0.70	0.00	0.24	0.00	0.70	0.00	0.40	0.00	0.70	0.00	0.24	0.00	0.70	0.00	0.40	0.00
x_7	0.50	0.00	0.54	0.25	0.50	0.00	0.54	0.41	0.50	0.00	0.56	0.24	0.50	0.00	0.56	0.42	0.50	0.00	0.56	0.24	0.50	0.00	0.56	0.42
x_8	0.60	0.00	0.00	0.25	0.60	0.00	0.00	0.38	0.60	0.00	0.00	0.24	0.60	0.00	0.00	0.36	0.60	0.00	0.00	0.24	0.60	0.00	0.00	0.36
x_9	0.35	0.00	0.00	0.38	0.35	0.00	0.00	0.47	0.35	0.00	0.00	0.36	0.35	0.00	0.00	0.45	0.35	0.00	0.00	0.36	0.35	0.00	0.00	0.45
x_{10}	0.25	0.00	0.00	0.44	0.25	0.00	0.00	0.51	0.25	0.00	0.00	0.42	0.25	0.00	0.00	0.48	0.25	0.00	0.00	0.42	0.25	0.00	0.00	0.48
Regression Slope																								
β	0.50	0.50	0.30	0.30	0.50	0.50	0.30	0.30	0.38	0.42	0.47	0.63	0.38	0.42	0.47	0.63	0.30	0.40	0.60	0.60	0.30	0.40	0.60	0.60
β^*	0.53	0.47	0.24	0.20	0.53	0.47	0.24	0.20	0.39	0.39	0.38	0.39	0.39	0.39	0.38	0.39	0.30	0.37	0.49	0.37	0.30	0.37	0.49	0.37
$var(\eta)$	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40	1.00	0.81	0.64	0.36	1.00	0.81	0.64	0.36	1.00	0.81	0.64	0.36	1.00	0.81	0.64	0.36
Factor Determinacy	0.91	0.62	0.72	0.61	0.90	0.79	0.76	0.75	0.91	0.62	0.72	0.61	0.90	0.79	0.76	0.75	0.91	0.62	0.72	0.61	0.90	0.79	0.76	0.75
$var(y)$	0.90				0.90				0.96				0.96				0.97				0.97			

Table 3
Summary of Classification and Spearman Rank Correlation for $g + 1s$

		g				g + 1 s			
		Sensitivity	Specificity	Correct	Rank Correlation	Sensitivity	Specificity	Correct	Rank Correlation
N=250 R=0.5									
Model	1	68.75%	68.75%	68.75%	0.539	69.62%	69.62%	69.62%	0.565
	2	70.20%	70.20%	70.20%	0.577	70.59%	70.59%	70.59%	0.588
	3	65.93%	65.93%	65.93%	0.462	68.47%	68.47%	68.47%	0.529
	4	68.13%	68.13%	68.13%	0.524	70.33%	70.33%	70.33%	0.578
	5	62.30%	62.30%	62.30%	0.364	66.56%	66.56%	66.56%	0.486
	6	65.66%	65.66%	65.66%	0.447	69.87%	69.87%	69.87%	0.566
	7	73.64%	73.64%	73.64%	0.660	75.04%	75.04%	75.04%	0.690
	8	75.61%	75.61%	75.61%	0.703	76.34%	76.34%	76.34%	0.720
	9	69.93%	69.93%	69.93%	0.568	73.47%	73.47%	73.47%	0.653
	10	73.04%	73.04%	73.04%	0.640	75.97%	75.97%	75.97%	0.708
	11	64.73%	64.73%	64.73%	0.435	71.22%	71.22%	71.22%	0.599
	12	68.87%	68.87%	68.87%	0.540	74.70%	74.70%	74.70%	0.686
N=250 R=0.1									
Model	1	35.45%	92.83%	87.09%	0.535	36.99%	93.00%	87.40%	0.558
	2	38.52%	93.17%	87.70%	0.573	39.31%	93.26%	87.86%	0.583
	3	30.94%	92.33%	86.19%	0.466	35.16%	92.80%	87.03%	0.528
	4	35.35%	92.82%	87.07%	0.525	38.48%	93.16%	87.70%	0.573
	5	24.98%	91.66%	85.00%	0.360	32.52%	92.50%	86.50%	0.486
	6	29.69%	92.19%	85.94%	0.439	36.85%	92.98%	87.37%	0.549
	7	45.67%	93.96%	89.13%	0.662	47.72%	94.19%	89.54%	0.692
	8	48.41%	94.27%	89.68%	0.707	49.78%	94.42%	89.96%	0.720
	9	38.18%	93.13%	87.64%	0.570	44.74%	93.86%	88.95%	0.653
	10	43.94%	93.77%	88.79%	0.645	49.22%	94.36%	89.84%	0.707
	11	30.23%	92.25%	86.05%	0.443	40.70%	93.41%	88.14%	0.600
	12	36.17%	92.91%	87.23%	0.543	47.51%	94.17%	89.50%	0.688
N=500 R=0.5									
Model	1	68.60%	68.60%	68.60%	0.535	69.56%	69.56%	69.56%	0.561
	2	70.08%	70.08%	70.08%	0.573	70.60%	70.60%	70.60%	0.586
	3	65.97%	65.97%	65.97%	0.461	68.52%	68.52%	68.52%	0.528
	4	68.44%	68.44%	68.44%	0.527	70.56%	70.56%	70.56%	0.579
	5	62.53%	62.53%	62.53%	0.363	67.12%	67.12%	67.12%	0.493
	6	65.18%	65.18%	65.18%	0.440	69.60%	69.60%	69.60%	0.558
	7	73.60%	73.60%	73.60%	0.659	75.12%	75.12%	75.12%	0.694
	8	75.91%	75.91%	75.91%	0.709	76.59%	76.59%	76.59%	0.725
	9	70.28%	70.28%	70.28%	0.575	73.55%	73.55%	73.55%	0.657
	10	73.12%	73.12%	73.12%	0.645	76.02%	76.02%	76.02%	0.711
	11	65.13%	65.13%	65.13%	0.446	71.48%	71.48%	71.48%	0.610
	12	68.86%	68.86%	68.86%	0.542	75.13%	75.13%	75.13%	0.692
N=500 R=0.1									
Model	1	35.68%	92.85%	87.14%	0.538	37.70%	93.08%	87.54%	0.563
	2	38.23%	93.14%	87.65%	0.575	39.25%	93.25%	87.85%	0.586
	3	31.54%	92.39%	86.31%	0.465	35.80%	92.87%	87.16%	0.532
	4	35.00%	92.78%	87.00%	0.524	38.42%	93.16%	87.68%	0.574
	5	25.52%	91.72%	85.10%	0.362	33.25%	92.58%	86.65%	0.491
	6	30.03%	92.23%	86.01%	0.443	37.23%	93.03%	87.45%	0.555
	7	45.23%	93.91%	89.05%	0.663	47.78%	94.20%	89.56%	0.695
	8	48.52%	94.28%	89.70%	0.707	49.72%	94.41%	89.94%	0.722
	9	37.95%	93.11%	87.59%	0.571	44.23%	93.80%	88.85%	0.655
	10	44.01%	93.78%	88.80%	0.648	49.21%	94.36%	89.84%	0.711
	11	29.86%	92.21%	85.97%	0.443	40.54%	93.39%	88.11%	0.605
	12	35.99%	92.89%	87.20%	0.541	47.57%	94.17%	89.51%	0.690

Note. N represents selected sample size.
R represents selection ratio.

Table 3

Summary of Classification and Spearman Rank Correlation for $g + 1s$

		Difference (gain from $g + 1s$)			
		Sensitivity	Specificity	Correct	Rank Correlation
N=250 R=0.5					
Model	1	0.87%	0.87%	0.87%	0.026
	2	0.39%	0.39%	0.39%	0.012
	3	2.54%	2.54%	2.54%	0.067
	4	2.20%	2.20%	2.20%	0.054
	5	4.26%	4.26%	4.26%	0.122
	6	4.21%	4.21%	4.21%	0.119
	7	1.41%	1.41%	1.41%	0.031
	8	0.73%	0.73%	0.73%	0.017
	9	3.54%	3.54%	3.54%	0.085
	10	2.92%	2.92%	2.92%	0.068
	11	6.48%	6.48%	6.48%	0.164
	12	5.84%	5.84%	5.84%	0.147
N=250 R=0.1					
Model	1	1.54%	0.17%	0.31%	0.022
	2	0.79%	0.09%	0.16%	0.010
	3	4.22%	0.47%	0.84%	0.062
	4	3.14%	0.35%	0.63%	0.048
	5	7.55%	0.84%	1.51%	0.126
	6	7.16%	0.80%	1.43%	0.111
	7	2.05%	0.23%	0.41%	0.030
	8	1.36%	0.15%	0.27%	0.013
	9	6.56%	0.73%	1.31%	0.083
	10	5.28%	0.59%	1.06%	0.062
	11	10.47%	1.16%	2.09%	0.157
	12	11.34%	1.26%	2.27%	0.145
N=500 R=0.5					
Model	1	0.96%	0.96%	0.96%	0.027
	2	0.52%	0.52%	0.52%	0.013
	3	2.55%	2.55%	2.55%	0.067
	4	2.13%	2.13%	2.13%	0.052
	5	4.60%	4.60%	4.60%	0.130
	6	4.42%	4.42%	4.42%	0.118
	7	1.51%	1.51%	1.51%	0.035
	8	0.68%	0.68%	0.68%	0.016
	9	3.27%	3.27%	3.27%	0.082
	10	2.90%	2.90%	2.90%	0.066
	11	6.35%	6.35%	6.35%	0.163
	12	6.27%	6.27%	6.27%	0.150
N=500 R=0.1					
Model	1	2.02%	0.22%	0.40%	0.025
	2	1.02%	0.11%	0.20%	0.011
	3	4.26%	0.47%	0.85%	0.067
	4	3.42%	0.38%	0.68%	0.050
	5	7.74%	0.86%	1.55%	0.129
	6	7.20%	0.80%	1.44%	0.113
	7	2.55%	0.28%	0.51%	0.032
	8	1.20%	0.13%	0.24%	0.015
	9	6.28%	0.70%	1.26%	0.085
	10	5.20%	0.58%	1.04%	0.064
	11	10.67%	1.19%	2.13%	0.162
	12	11.58%	1.29%	2.32%	0.149

Table 4

Summary of Classification and Spearman Rank Correlation for $g + 2s$

		g				$g + 2s$			
		Sensitivity	Specificity	Correct	Rank Correlation	Sensitivity	Specificity	Correct	Rank Correlation
N=250 R=0.5									
Model	1	69.13%	69.13%	69.13%	0.548	69.40%	69.40%	69.40%	0.556
	2	70.10%	70.10%	70.10%	0.570	70.83%	70.83%	70.83%	0.591
	3	65.65%	65.65%	65.65%	0.452	66.57%	66.57%	66.57%	0.475
	4	66.72%	66.72%	66.72%	0.490	68.64%	68.64%	68.64%	0.536
	5	64.61%	64.61%	64.61%	0.426	65.56%	65.56%	65.56%	0.453
	6	66.25%	66.25%	66.25%	0.471	68.27%	68.27%	68.27%	0.522
	7	74.38%	74.38%	74.38%	0.674	74.68%	74.68%	74.68%	0.684
	8	74.94%	74.94%	74.94%	0.694	76.08%	76.08%	76.08%	0.718
	9	68.48%	68.48%	68.48%	0.532	70.56%	70.56%	70.56%	0.564
	10	71.23%	71.23%	71.23%	0.605	73.50%	73.50%	73.50%	0.661
	11	68.56%	68.56%	68.56%	0.526	69.72%	69.72%	69.72%	0.561
	12	70.31%	70.31%	70.31%	0.576	72.90%	72.90%	72.90%	0.641
N=250 R=0.1									
Model	1	36.90%	92.99%	87.38%	0.548	37.56%	93.06%	87.51%	0.553
	2	38.05%	93.12%	87.61%	0.568	39.34%	93.26%	87.87%	0.583
	3	30.87%	92.32%	86.17%	0.454	32.09%	92.45%	86.42%	0.473
	4	33.38%	92.60%	86.68%	0.492	36.26%	92.92%	87.25%	0.536
	5	29.27%	92.14%	85.85%	0.429	30.96%	92.33%	86.19%	0.454
	6	31.50%	92.39%	86.30%	0.472	34.46%	92.72%	86.89%	0.520
	7	45.63%	93.96%	89.13%	0.676	46.79%	[illegible]	89.36%	0.685
	8	47.88%	94.21%	89.58%	0.702	50.39%	94.49%	90.08%	0.725
	9	37.33%	93.04%	87.47%	0.559	39.71%	93.30%	87.94%	0.591
	10	40.57%	93.40%	88.11%	0.604	45.10%	93.90%	89.02%	0.664
	11	35.52%	92.84%	87.10%	0.526	38.13%	93.13%	87.63%	0.564
	12	38.70%	93.19%	87.74%	0.580	43.83%	93.76%	88.77%	0.647
N=500 R=0.5									
Model	1	69.08%	69.08%	69.08%	0.546	69.42%	69.42%	69.42%	0.557
	2	70.07%	70.07%	70.07%	0.567	70.81%	70.81%	70.81%	0.586
	3	65.36%	65.36%	65.36%	0.452	66.44%	66.44%	66.44%	0.479
	4	66.82%	66.82%	66.82%	0.488	68.61%	68.61%	68.61%	0.538
	5	64.72%	64.72%	64.72%	0.429	65.99%	65.99%	65.99%	0.462
	6	66.05%	66.05%	66.05%	0.467	67.86%	67.86%	67.86%	0.521
	7	74.43%	74.43%	74.43%	0.677	75.03%	75.03%	75.03%	0.689
	8	75.44%	75.44%	75.44%	0.701	76.79%	76.79%	76.79%	0.727
	9	69.63%	69.63%	69.63%	0.560	70.76%	70.76%	70.76%	0.592
	10	71.43%	71.43%	71.43%	0.606	74.04%	74.04%	74.04%	0.666
	11	68.24%	68.24%	68.24%	0.524	69.85%	69.85%	69.85%	0.565
	12	70.49%	70.49%	70.49%	0.575	73.23%	73.23%	73.23%	0.645
N=500 R=0.1									
Model	1	36.72%	92.97%	87.34%	0.549	37.27%	93.03%	87.45%	0.557
	2	38.11%	93.12%	87.62%	0.568	39.57%	93.29%	87.91%	0.586
	3	30.69%	92.30%	86.14%	0.454	32.30%	92.48%	86.46%	0.479
	4	32.98%	92.55%	86.60%	0.487	36.32%	92.92%	87.26%	0.536
	5	29.17%	92.13%	85.83%	0.425	31.24%	92.36%	86.25%	0.458
	6	31.49%	92.39%	86.30%	0.471	35.39%	92.82%	87.08%	0.526
	7	46.20%	94.02%	89.24%	0.675	47.12%	94.12%	89.42%	0.686
	8	48.44%	94.27%	89.69%	0.701	50.68%	94.52%	90.14%	0.727
	9	37.58%	93.06%	87.52%	0.562	39.96%	93.33%	87.99%	0.596
	10	40.38%	93.38%	88.08%	0.605	45.38%	93.93%	89.08%	0.667
	11	35.47%	92.83%	87.09%	0.525	38.37%	93.15%	87.67%	0.568
	12	38.70%	93.19%	87.74%	0.577	44.12%	93.79%	88.82%	0.650

Note. N represents selected sample size.
R represents selection ratio.

Table 4

Summary of Classification and Spearman Rank Correlation for $g + 2s$

		Difference (gain from $g + 2s$)			
		Sensitivity	Specificity	Correct	Rank Correlation
N=250 R=0.5					
Model	1	0.27%	0.27%	0.27%	0.007
	2	0.73%	0.73%	0.73%	0.020
	3	0.92%	0.92%	0.92%	0.023
	4	1.93%	1.93%	1.93%	0.046
	5	0.94%	0.94%	0.94%	0.027
	6	2.02%	2.02%	2.02%	0.052
	7	0.29%	0.29%	0.29%	0.010
	8	1.14%	1.14%	1.14%	0.024
	9	2.08%	2.08%	2.08%	0.032
	10	2.27%	2.27%	2.27%	0.056
	11	1.16%	1.16%	1.16%	0.035
	12	2.58%	2.58%	2.58%	0.065
N=250 R=0.1					
Model	1	0.66%	0.07%	0.13%	0.005
	2	1.29%	0.14%	0.26%	0.015
	3	1.22%	0.14%	0.24%	0.020
	4	2.88%	0.32%	0.58%	0.044
	5	1.69%	0.19%	0.34%	0.025
	6	2.95%	0.33%	0.59%	0.048
	7	1.16%	0.13%	0.23%	0.010
	8	2.50%	0.28%	0.50%	0.024
	9	2.38%	0.26%	0.48%	0.032
	10	4.53%	0.50%	0.91%	0.059
	11	2.61%	0.29%	0.52%	0.037
	12	5.13%	0.57%	1.03%	0.067
N=500 R=0.5					
Model	1	0.34%	0.34%	0.34%	0.010
	2	0.74%	0.74%	0.74%	0.019
	3	1.09%	1.09%	1.09%	0.027
	4	1.80%	1.80%	1.80%	0.050
	5	1.27%	1.27%	1.27%	0.032
	6	1.81%	1.81%	1.81%	0.054
	7	0.60%	0.60%	0.60%	0.012
	8	1.35%	1.35%	1.35%	0.026
	9	1.13%	1.13%	1.13%	0.031
	10	2.62%	2.62%	2.62%	0.061
	11	1.61%	1.61%	1.61%	0.041
	12	2.75%	2.75%	2.75%	0.070
N=500 R=0.1					
Model	1	0.56%	0.06%	0.11%	0.008
	2	1.46%	0.16%	0.29%	0.019
	3	1.60%	0.18%	0.32%	0.025
	4	3.33%	0.37%	0.67%	0.049
	5	2.07%	0.23%	0.41%	0.033
	6	3.90%	0.43%	0.78%	0.055
	7	0.92%	0.10%	0.18%	0.011
	8	2.24%	0.25%	0.45%	0.025
	9	2.39%	0.27%	0.48%	0.034
	10	5.00%	0.55%	1.00%	0.062
	11	2.90%	0.32%	0.58%	0.043
	12	5.43%	0.60%	1.09%	0.072

Table 5

Summary of Classification and Spearman Rank Correlation for $g + 3s$

		g				$g + 3s$			
		Sensitivity	Specificity	Correct	Rank Correlation	Sensitivity	Specificity	Correct	Rank Correlation
N=250 R=0.5									
Model	1	69.09%	69.09%	69.09%	0.547	69.31%	69.31%	69.31%	0.556
	2	69.72%	69.72%	69.72%	0.564	70.32%	70.32%	70.32%	0.582
	3	65.73%	65.73%	65.73%	0.458	67.36%	67.36%	67.36%	0.501
	4	66.73%	66.73%	66.73%	0.479	69.68%	69.68%	69.68%	0.554
	5	64.30%	64.30%	64.30%	0.412	66.35%	66.35%	66.35%	0.479
	6	64.64%	64.64%	64.64%	0.423	68.44%	68.44%	68.44%	0.529
	7	74.22%	74.22%	74.22%	0.670	74.66%	74.66%	74.66%	0.683
	8	75.02%	75.02%	75.02%	0.696	76.16%	76.16%	76.16%	0.718
	9	69.76%	69.76%	69.76%	0.563	71.68%	71.68%	71.68%	0.617
	10	70.59%	70.59%	70.59%	0.580	74.29%	74.29%	74.29%	0.673
	11	67.20%	67.20%	67.20%	0.503	70.67%	70.67%	70.67%	0.593
	12	68.17%	68.17%	68.17%	0.519	73.29%	73.29%	73.29%	0.650
N=250 R=0.1									
Model	1	36.42%	92.94%	87.28%	0.544	36.98%	93.00%	87.40%	0.546
	2	37.76%	93.08%	87.55%	0.566	38.68%	93.19%	87.74%	0.576
	3	30.81%	92.31%	86.16%	0.455	33.15%	92.57%	86.63%	0.489
	4	31.74%	92.42%	86.35%	0.469	36.09%	92.90%	87.22%	0.534
	5	28.27%	92.03%	85.65%	0.411	32.02%	92.45%	86.40%	0.472
	6	28.64%	92.07%	85.73%	0.423	34.69%	92.74%	86.94%	0.520
	7	45.72%	93.97%	89.14%	0.674	46.57%	94.06%	89.31%	0.682
	8	48.14%	94.24%	89.63%	0.700	49.77%	94.42%	89.95%	0.718
	9	37.40%	93.04%	87.48%	0.562	41.44%	93.49%	88.29%	0.614
	10	38.74%	93.19%	87.75%	0.578	45.68%	93.96%	89.14%	0.668
	11	33.94%	92.66%	86.79%	0.503	39.72%	93.30%	87.94%	0.587
	12	34.47%	92.72%	86.89%	0.521	44.01%	93.78%	88.80%	0.646
N=500 R=0.5									
Model	1	69.09%	69.09%	69.09%	0.545	69.63%	69.63%	69.63%	0.556
	2	69.92%	69.92%	69.92%	0.566	70.56%	70.56%	70.56%	0.584
	3	65.68%	65.68%	65.68%	0.453	67.40%	67.40%	67.40%	0.500
	4	66.02%	66.02%	66.02%	0.469	68.90%	68.90%	68.90%	0.544
	5	64.15%	64.15%	64.15%	0.411	66.74%	66.74%	66.74%	0.483
	6	64.50%	64.50%	64.50%	0.426	68.54%	68.54%	68.54%	0.532
	7	74.26%	74.26%	74.26%	0.672	74.88%	74.88%	74.88%	0.684
	8	75.60%	75.60%	75.60%	0.700	76.56%	76.56%	76.56%	0.721
	9	69.66%	69.66%	69.66%	0.557	71.88%	71.88%	71.88%	0.615
	10	70.37%	70.37%	70.37%	0.582	74.21%	74.21%	74.21%	0.675
	11	67.49%	67.49%	67.49%	0.504	71.10%	71.10%	71.10%	0.596
	12	68.20%	68.20%	68.20%	0.521	73.52%	73.52%	73.52%	0.653
N=500 R=0.1									
Model	1	36.35%	92.93%	87.27%	0.546	37.05%	93.01%	87.41%	0.553
	2	38.24%	93.14%	87.65%	0.566	39.43%	93.27%	87.89%	0.581
	3	30.75%	92.31%	86.15%	0.456	33.26%	92.58%	86.65%	0.497
	4	31.49%	92.39%	86.30%	0.470	36.28%	92.92%	87.26%	0.542
	5	28.20%	92.02%	85.64%	0.410	32.63%	92.51%	86.53%	0.480
	6	29.02%	92.11%	85.80%	0.423	35.21%	92.80%	87.04%	0.526
	7	45.74%	93.97%	89.15%	0.675	46.91%	94.10%	89.38%	0.686
	8	48.30%	94.26%	89.66%	0.701	50.15%	94.46%	90.03%	0.721
	9	37.67%	93.07%	87.53%	0.561	41.77%	93.53%	88.35%	0.617
	10	38.88%	93.21%	87.78%	0.580	46.02%	94.00%	89.20%	0.672
	11	33.43%	92.60%	86.69%	0.503	39.77%	93.31%	87.95%	0.593
	12	34.97%	92.77%	86.99%	0.520	44.39%	93.82%	88.88%	0.651

Note. N represents selected sample size.
R represents selection ratio.

Table 5

Summary of Classification and Spearman Rank Correlation for $g + 3s$

		Difference (gain from $g + 3s$)			
		Sensitivity	Specificity	Correct	Rank Correlation
N=250	R=0.5				
Model	1	0.22%	0.22%	0.22%	0.009
	2	0.60%	0.60%	0.60%	0.017
	3	1.63%	1.63%	1.63%	0.043
	4	2.94%	2.94%	2.94%	0.075
	5	2.05%	2.05%	2.05%	0.067
	6	3.80%	3.80%	3.80%	0.106
	7	0.44%	0.44%	0.44%	0.012
	8	1.14%	1.14%	1.14%	0.022
	9	1.92%	1.92%	1.92%	0.053
	10	3.70%	3.70%	3.70%	0.093
	11	3.47%	3.47%	3.47%	0.090
	12	5.12%	5.12%	5.12%	0.131
N=250	R=0.1				
Model	1	0.56%	0.06%	0.11%	0.002
	2	0.92%	0.10%	0.18%	0.010
	3	2.34%	0.26%	0.47%	0.034
	4	4.34%	0.48%	0.87%	0.065
	5	3.75%	0.42%	0.75%	0.061
	6	6.05%	0.67%	1.21%	0.097
	7	0.85%	0.09%	0.17%	0.009
	8	1.64%	0.18%	0.33%	0.018
	9	4.04%	0.45%	0.81%	0.052
	10	6.94%	0.77%	1.39%	0.090
	11	5.78%	0.64%	1.16%	0.084
	12	9.54%	1.06%	1.91%	0.125
N=500	R=0.5				
Model	1	0.55%	0.55%	0.55%	0.011
	2	0.64%	0.64%	0.64%	0.017
	3	1.71%	1.71%	1.71%	0.047
	4	2.88%	2.88%	2.88%	0.076
	5	2.59%	2.59%	2.59%	0.072
	6	4.04%	4.04%	4.04%	0.106
	7	0.62%	0.62%	0.62%	0.012
	8	0.96%	0.96%	0.96%	0.021
	9	2.22%	2.22%	2.22%	0.058
	10	3.84%	3.84%	3.84%	0.092
	11	3.61%	3.61%	3.61%	0.091
	12	5.32%	5.32%	5.32%	0.132
N=500	R=0.1				
Model	1	0.70%	0.08%	0.14%	0.007
	2	1.20%	0.13%	0.24%	0.014
	3	2.51%	0.28%	0.50%	0.041
	4	4.79%	0.53%	0.96%	0.072
	5	4.43%	0.49%	0.89%	0.070
	6	6.18%	0.69%	1.24%	0.103
	7	1.16%	0.13%	0.23%	0.011
	8	1.85%	0.21%	0.37%	0.020
	9	4.10%	0.46%	0.82%	0.056
	10	7.14%	0.79%	1.43%	0.092
	11	6.34%	0.70%	1.27%	0.090
	12	9.42%	1.05%	1.88%	0.131

Table 6
Summary of Cross-Validation for $g + 1s$

	Spearman Rank Correlation						PRESS					
	Sample 1			Sample 2			Sample 1			Sample 2		
	$g+1s$	g	Difference	$g+1s$	g	Difference	$g+1s$	g	Difference	$g+1s$	g	Difference
N=250 R=0.5												
Model 1	0.389	0.353	0.036	0.398	0.355	0.043	0.907	0.921	-0.014	0.900	0.917	-0.017
2	0.406	0.392	0.013	0.391	0.378	0.014	0.890	0.895	-0.005	0.889	0.894	-0.005
3	0.403	0.305	0.098	0.397	0.293	0.104	0.924	0.962	-0.038	0.927	0.969	-0.042
4	0.437	0.338	0.099	0.416	0.336	0.079	0.883	0.923	-0.040	0.894	0.929	-0.035
5	0.403	0.223	0.180	0.406	0.229	0.177	0.952	1.016	-0.065	0.965	1.030	-0.065
6	0.469	0.291	0.179	0.456	0.275	0.181	0.912	0.998	-0.086	0.902	0.980	-0.078
7	0.518	0.462	0.056	0.508	0.454	0.055	0.642	0.664	-0.022	0.642	0.666	-0.023
8	0.526	0.495	0.031	0.523	0.490	0.033	0.620	0.633	-0.013	0.615	0.630	-0.015
9	0.522	0.373	0.149	0.521	0.376	0.145	0.671	0.733	-0.062	0.675	0.732	-0.057
10	0.565	0.437	0.129	0.551	0.433	0.118	0.625	0.682	-0.057	0.635	0.686	-0.051
11	0.517	0.274	0.243	0.515	0.271	0.244	0.711	0.806	-0.095	0.716	0.811	-0.094
12	0.578	0.347	0.232	0.581	0.352	0.229	0.645	0.752	-0.107	0.638	0.746	-0.108
N=250 R=0.1												
Model 1	0.286	0.226	0.060	0.285	0.221	0.064	0.918	0.933	-0.015	0.901	0.918	-0.017
2	0.281	0.250	0.031	0.269	0.247	0.023	0.889	0.895	-0.006	0.895	0.900	-0.005
3	0.340	0.192	0.148	0.346	0.200	0.146	0.926	0.967	-0.041	0.929	0.974	-0.045
4	0.341	0.228	0.113	0.357	0.234	0.123	0.899	0.933	-0.035	0.888	0.928	-0.040
5	0.361	0.144	0.217	0.374	0.132	0.242	0.956	1.020	-0.065	0.952	1.019	-0.067
6	0.412	0.178	0.234	0.417	0.176	0.241	0.910	0.989	-0.079	0.904	0.981	-0.077
7	0.400	0.311	0.089	0.398	0.302	0.097	0.642	0.667	-0.025	0.638	0.663	-0.025
8	0.401	0.357	0.044	0.397	0.353	0.045	0.615	0.626	-0.012	0.611	0.623	-0.012
9	0.467	0.273	0.194	0.452	0.251	0.201	0.670	0.733	-0.062	0.674	0.734	-0.060
10	0.488	0.308	0.180	0.468	0.300	0.168	0.619	0.674	-0.055	0.617	0.669	-0.052
11	0.489	0.198	0.291	0.488	0.157	0.332	0.714	0.809	-0.094	0.705	0.805	-0.100
12	0.545	0.250	0.295	0.531	0.236	0.295	0.634	0.739	-0.105	0.646	0.748	-0.102
N=500 R=0.5												
Model 1	0.403	0.358	0.045	0.398	0.350	0.048	0.899	0.906	-0.007	0.905	0.923	-0.018
2	0.399	0.376	0.022	0.394	0.373	0.021	0.885	0.894	-0.010	0.884	0.892	-0.008
3	0.398	0.282	0.115	0.399	0.292	0.106	0.929	0.974	-0.046	0.922	0.963	-0.041
4	0.429	0.343	0.086	0.429	0.342	0.088	0.886	0.924	-0.038	0.883	0.922	-0.039
5	0.412	0.220	0.193	0.418	0.222	0.196	0.950	1.024	-0.074	0.944	1.018	-0.075
6	0.459	0.274	0.186	0.455	0.276	0.179	0.901	0.980	-0.079	0.905	0.981	-0.076
7	0.516	0.444	0.072	0.518	0.447	0.071	0.639	0.668	-0.029	0.639	0.669	-0.030
8	0.532	0.498	0.033	0.537	0.508	0.029	0.607	0.621	-0.014	0.605	0.618	-0.013
9	0.520	0.377	0.143	0.527	0.380	0.148	0.671	0.731	-0.059	0.665	0.727	-0.062
10	0.565	0.445	0.121	0.562	0.438	0.124	0.620	0.676	-0.056	0.622	0.678	-0.055
11	0.528	0.292	0.236	0.524	0.279	0.245	0.705	0.802	-0.097	0.701	0.799	-0.097
12	0.597	0.357	0.240	0.591	0.351	0.240	0.633	0.743	-0.110	0.636	0.747	-0.111
N=500 R=0.1												
Model 1	0.306	0.245	0.061	0.323	0.248	0.075	0.897	0.914	-0.018	0.897	0.919	-0.021
2	0.284	0.249	0.034	0.283	0.262	0.021	0.885	0.894	-0.009	0.879	0.885	-0.006
3	0.354	0.204	0.150	0.358	0.204	0.154	0.923	0.967	-0.044	0.929	0.977	-0.048
4	0.355	0.218	0.137	0.357	0.232	0.124	0.891	0.930	-0.039	0.891	0.927	-0.037
5	0.387	0.141	0.246	0.387	0.143	0.244	0.948	1.019	-0.072	0.952	1.026	-0.074
6	0.420	0.176	0.244	0.419	0.179	0.240	0.905	0.987	-0.082	0.907	0.988	-0.081
7	0.410	0.318	0.092	0.412	0.307	0.105	0.638	0.663	-0.025	0.633	0.661	-0.028
8	0.408	0.356	0.052	0.405	0.355	0.051	0.613	0.626	-0.014	0.610	0.623	-0.013
9	0.463	0.263	0.200	0.457	0.265	0.192	0.665	0.726	-0.062	0.666	0.726	-0.060
10	0.483	0.312	0.171	0.475	0.301	0.174	0.617	0.672	-0.055	0.619	0.674	-0.055
11	0.496	0.185	0.311	0.493	0.179	0.314	0.701	0.802	-0.100	0.701	0.801	-0.100
12	0.551	0.245	0.306	0.555	0.241	0.314	0.641	0.753	-0.111	0.636	0.750	-0.113

Note. N represents selected sample size.
R represents selection ratio.

Table 7
Summary of Cross-Validation for $g + 2s$

		Spearman Rank Correlation						PRESS					
		Sample 1			Sample 2			Sample 1			Sample 2		
		$g+2s$	g	Difference	$g+2s$	g	Difference	$g+2s$	g	Difference	$g+2s$	g	Difference
N=250 R=0.5													
Model	1	0.354	0.352	0.002	0.352	0.354	-0.002	0.683	0.683	0.001	0.677	0.678	-0.001
	2	0.403	0.377	0.026	0.390	0.359	0.031	0.655	0.661	-0.006	0.662	0.670	-0.007
	3	0.311	0.290	0.021	0.314	0.281	0.033	0.705	0.710	-0.005	0.701	0.709	-0.008
	4	0.379	0.315	0.064	0.375	0.308	0.067	0.672	0.689	-0.017	0.677	0.695	-0.018
	5	0.300	0.254	0.045	0.292	0.265	0.027	0.836	0.847	-0.011	0.835	0.843	-0.008
	6	0.377	0.297	0.080	0.370	0.290	0.080	0.797	0.822	-0.026	0.790	0.814	-0.024
	7	0.477	0.468	0.009	0.477	0.461	0.016	0.485	0.488	-0.003	0.480	0.484	-0.003
	8	0.533	0.492	0.040	0.531	0.489	0.042	0.455	0.468	-0.013	0.458	0.473	-0.014
	9	0.421	0.397	0.024	0.389	0.384	0.005	0.512	0.512	0.000	0.536	0.535	0.002
	10	0.503	0.409	0.093	0.497	0.403	0.094	0.491	0.519	-0.027	0.488	0.517	-0.029
	11	0.383	0.341	0.042	0.381	0.331	0.051	0.632	0.643	-0.011	0.634	0.648	-0.015
	12	0.485	0.368	0.118	0.483	0.379	0.104	0.581	0.623	-0.042	0.581	0.617	-0.036
N=250 R=0.1													
Model	1	0.235	0.238	-0.003	0.246	0.239	0.007	0.687	0.687	0.000	0.679	0.679	0.000
	2	0.294	0.257	0.036	0.304	0.257	0.047	0.663	0.669	-0.006	0.665	0.675	-0.010
	3	0.245	0.179	0.066	0.256	0.194	0.062	0.702	0.709	-0.007	0.698	0.708	-0.010
	4	0.321	0.212	0.110	0.310	0.200	0.110	0.676	0.697	-0.021	0.667	0.686	-0.020
	5	0.248	0.181	0.068	0.222	0.148	0.074	0.833	0.845	-0.012	0.843	0.853	-0.010
	6	0.325	0.205	0.120	0.311	0.193	0.118	0.789	0.815	-0.026	0.795	0.820	-0.025
	7	0.356	0.333	0.023	0.353	0.329	0.024	0.482	0.486	-0.003	0.480	0.483	-0.004
	8	0.421	0.352	0.070	0.415	0.342	0.073	0.453	0.467	-0.014	0.455	0.470	-0.015
	9	0.342	0.256	0.086	0.339	0.239	0.100	0.515	0.529	-0.015	0.521	0.536	-0.016
	10	0.425	0.268	0.158	0.421	0.264	0.158	0.482	0.515	-0.033	0.481	0.513	-0.032
	11	0.331	0.243	0.088	0.326	0.230	0.097	0.623	0.641	-0.018	0.628	0.648	-0.020
	12	0.427	0.272	0.155	0.425	0.259	0.166	0.577	0.616	-0.040	0.577	0.615	-0.039
N=500 R=0.5													
Model	1	0.371	0.355	0.016	0.369	0.352	0.017	0.669	0.672	-0.004	0.672	0.677	-0.004
	2	0.401	0.374	0.028	0.402	0.370	0.032	0.651	0.660	-0.008	0.649	0.659	-0.010
	3	0.328	0.287	0.041	0.330	0.289	0.041	0.693	0.702	-0.009	0.692	0.701	-0.010
	4	0.394	0.304	0.089	0.390	0.308	0.082	0.667	0.691	-0.025	0.665	0.688	-0.023
	5	0.313	0.264	0.048	0.318	0.278	0.041	0.828	0.841	-0.012	0.827	0.839	-0.012
	6	0.388	0.298	0.091	0.390	0.300	0.090	0.787	0.815	-0.029	0.787	0.816	-0.029
	7	0.488	0.467	0.021	0.494	0.474	0.020	0.479	0.485	-0.007	0.479	0.485	-0.006
	8	0.543	0.490	0.053	0.545	0.496	0.049	0.450	0.467	-0.017	0.451	0.469	-0.017
	9	0.422	0.370	0.053	0.415	0.365	0.050	0.525	0.539	-0.014	0.517	0.530	-0.012
	10	0.507	0.402	0.105	0.507	0.399	0.108	0.482	0.513	-0.031	0.481	0.512	-0.032
	11	0.405	0.339	0.066	0.405	0.338	0.067	0.623	0.642	-0.019	0.619	0.638	-0.019
	12	0.495	0.373	0.122	0.494	0.369	0.125	0.571	0.613	-0.042	0.569	0.610	-0.042
N=500 R=0.1													
Model	1	0.265	0.241	0.024	0.266	0.243	0.023	0.681	0.685	-0.004	0.671	0.675	-0.004
	2	0.303	0.253	0.049	0.309	0.252	0.058	0.662	0.671	-0.009	0.654	0.666	-0.012
	3	0.253	0.195	0.058	0.259	0.198	0.061	0.699	0.707	-0.008	0.694	0.703	-0.010
	4	0.327	0.210	0.116	0.320	0.211	0.109	0.666	0.690	-0.024	0.668	0.689	-0.021
	5	0.267	0.189	0.078	0.259	0.177	0.082	0.827	0.842	-0.015	0.821	0.836	-0.015
	6	0.335	0.204	0.131	0.327	0.208	0.118	0.789	0.820	-0.031	0.786	0.816	-0.029
	7	0.359	0.325	0.034	0.363	0.326	0.037	0.477	0.483	-0.006	0.478	0.485	-0.007
	8	0.434	0.352	0.082	0.421	0.347	0.075	0.449	0.466	-0.017	0.453	0.467	-0.014
	9	0.348	0.252	0.096	0.343	0.243	0.099	0.520	0.537	-0.016	0.514	0.530	-0.016
	10	0.441	0.285	0.156	0.443	0.277	0.166	0.477	0.512	-0.035	0.473	0.509	-0.036
	11	0.333	0.231	0.103	0.332	0.212	0.120	0.622	0.643	-0.021	0.621	0.645	-0.024
	12	0.446	0.261	0.185	0.445	0.265	0.180	0.568	0.614	-0.046	0.566	0.612	-0.045

Note. N represents selected sample size.
R represents selection ratio.

Table 8
Summary of Cross-Validation for $g + 3s$

		Spearman Rank Correlation						PRESS					
		Sample 1			Sample 2			Sample 1			Sample 2		
		$g+3s$	g	Difference	$g+3s$	g	Difference	$g+3s$	g	Difference	$g+3s$	g	Difference
N=250	R=0.5												
Model	1	0.352	0.359	-0.007	0.364	0.372	-0.008	0.969	0.966	0.003	0.959	0.953	0.006
	2	0.388	0.379	0.009	0.382	0.365	0.017	0.940	0.942	-0.002	0.943	0.949	-0.006
	3	0.342	0.285	0.057	0.339	0.294	0.045	1.045	1.068	-0.023	1.047	1.065	-0.019
	4	0.415	0.307	0.108	0.399	0.305	0.094	1.017	1.062	-0.045	1.013	1.054	-0.041
	5	0.331	0.241	0.091	0.343	0.246	0.097	1.058	1.086	-0.028	1.054	1.092	-0.038
	6	0.407	0.275	0.133	0.406	0.260	0.145	1.020	1.075	-0.056	1.019	1.081	-0.062
	7	0.474	0.463	0.011	0.465	0.461	0.004	0.684	0.688	-0.004	0.696	0.697	0.000
	8	0.528	0.496	0.032	0.519	0.492	0.027	0.652	0.666	-0.014	0.656	0.667	-0.011
	9	0.451	0.381	0.071	0.430	0.350	0.080	0.772	0.797	-0.025	0.766	0.793	-0.028
	10	0.528	0.380	0.148	0.533	0.387	0.146	0.713	0.781	-0.067	0.720	0.791	-0.071
	11	0.470	0.328	0.143	0.459	0.332	0.127	0.786	0.845	-0.059	0.784	0.837	-0.052
	12	0.519	0.340	0.179	0.523	0.308	0.214	0.744	0.828	-0.084	0.747	0.842	-0.095
N=250	R=0.1												
Model	1	0.232	0.239	-0.007	0.241	0.243	-0.002	0.972	0.968	0.005	0.977	0.974	0.004
	2	0.282	0.265	0.017	0.272	0.241	0.031	0.942	0.946	-0.004	0.955	0.961	-0.006
	3	0.268	0.185	0.083	0.263	0.178	0.085	1.048	1.067	-0.019	1.046	1.064	-0.018
	4	0.339	0.190	0.149	0.335	0.187	0.149	1.010	1.055	-0.044	1.002	1.046	-0.044
	5	0.300	0.170	0.130	0.296	0.162	0.134	1.061	1.095	-0.035	1.063	1.097	-0.033
	6	0.352	0.166	0.186	0.353	0.168	0.185	1.018	1.074	-0.056	1.019	1.074	-0.056
	7	0.339	0.329	0.010	0.333	0.315	0.018	0.687	0.689	-0.001	0.691	0.695	-0.003
	8	0.398	0.340	0.058	0.395	0.339	0.056	0.652	0.666	-0.014	0.647	0.662	-0.015
	9	0.385	0.262	0.123	0.370	0.255	0.115	0.771	0.808	-0.036	0.766	0.801	-0.035
	10	0.486	0.269	0.218	0.455	0.251	0.203	0.712	0.786	-0.074	0.724	0.787	-0.063
	11	0.412	0.226	0.186	0.393	0.193	0.200	0.782	0.838	-0.056	0.782	0.834	-0.052
	12	0.497	0.228	0.268	0.487	0.222	0.264	0.743	0.837	-0.094	0.741	0.833	-0.092
N=500	R=0.5												
Model	1	0.365	0.364	0.001	0.359	0.350	0.009	0.958	0.958	0.000	0.960	0.964	-0.004
	2	0.392	0.370	0.022	0.391	0.371	0.020	0.933	0.943	-0.010	0.940	0.948	-0.009
	3	0.350	0.284	0.067	0.359	0.286	0.073	1.034	1.058	-0.024	1.031	1.060	-0.029
	4	0.413	0.295	0.118	0.401	0.287	0.114	0.999	1.052	-0.053	0.997	1.047	-0.050
	5	0.359	0.254	0.104	0.360	0.242	0.118	1.056	1.095	-0.039	1.049	1.091	-0.042
	6	0.419	0.264	0.155	0.423	0.263	0.160	1.008	1.075	-0.067	1.011	1.082	-0.071
	7	0.487	0.467	0.020	0.489	0.473	0.017	0.680	0.689	-0.009	0.679	0.686	-0.008
	8	0.532	0.493	0.040	0.526	0.492	0.034	0.647	0.666	-0.019	0.642	0.659	-0.016
	9	0.465	0.366	0.099	0.452	0.355	0.097	0.755	0.795	-0.040	0.766	0.804	-0.038
	10	0.539	0.382	0.157	0.548	0.391	0.157	0.711	0.781	-0.070	0.709	0.786	-0.076
	11	0.466	0.325	0.141	0.465	0.323	0.142	0.785	0.841	-0.056	0.785	0.840	-0.055
	12	0.542	0.329	0.213	0.529	0.330	0.200	0.737	0.834	-0.097	0.741	0.832	-0.090
N=500	R=0.1												
Model	1	0.264	0.250	0.014	0.263	0.243	0.020	0.955	0.958	-0.003	0.952	0.956	-0.004
	2	0.296	0.256	0.040	0.303	0.260	0.043	0.937	0.946	-0.009	0.941	0.953	-0.012
	3	0.292	0.183	0.109	0.293	0.196	0.098	1.027	1.058	-0.030	1.028	1.057	-0.029
	4	0.366	0.201	0.165	0.361	0.202	0.159	0.997	1.052	-0.054	1.000	1.054	-0.054
	5	0.315	0.170	0.145	0.314	0.166	0.149	1.048	1.090	-0.042	1.050	1.093	-0.043
	6	0.377	0.180	0.197	0.381	0.173	0.208	1.018	1.083	-0.066	1.015	1.086	-0.071
	7	0.353	0.319	0.035	0.364	0.322	0.042	0.676	0.684	-0.008	0.674	0.684	-0.010
	8	0.417	0.349	0.068	0.412	0.353	0.059	0.644	0.663	-0.020	0.643	0.660	-0.017
	9	0.395	0.253	0.142	0.390	0.251	0.140	0.763	0.806	-0.043	0.759	0.800	-0.041
	10	0.484	0.265	0.220	0.485	0.262	0.223	0.711	0.786	-0.075	0.707	0.783	-0.077
	11	0.415	0.217	0.198	0.416	0.228	0.188	0.778	0.839	-0.060	0.785	0.843	-0.058
	12	0.508	0.235	0.274	0.502	0.220	0.282	0.730	0.831	-0.101	0.728	0.827	-0.099

Note. N represents selected sample size.
R represents selection ratio.

Table 9
The Number of Incumbents in the Army Project A and the Marine Corps JPM Studies

Enlisted Job	N
Army Project A	
1. Infantryman (11B)	491
2. Cannon Crewman (13B)	464
3. Tank Crewman (19E)	394
4. Radio Operator (31C)	289
5. Vehicle Mechanic (63B)	478
6. Motor Transport (64C)	507
7. Administrative (71L)	427
8. Medical (91A)	392
9. Military Police (95B)	597
Total	4,039
Marine Corps JPM	
10. Helicopter Mechanic	439
11. Automotive Mechanic	694
Total	1,133

Table 10
Summary of Tests in the Studies

Subtest	Description	Number of Items
ASVAB subtests		
GS	General Science	25
AR	Arithmetic Reasoning	30
WK	Word Knowledge	35
PC	Paragraph Comprehension	15
NO	Numerical Operations	50
CS	Coding Speed	84
AS	Auto & Shop Information	25
MK	Mathematical Knowledge	25
MC	Mechanical Comprehension	25
EI	Electronics Information	20
Army Project A subtests		
ASSEM.OBJ	Assembling Objects	32
REASON	Figural Reasoning	35
MAZE	Maze Test	24
OBJ.ROT	Object Rotation Test	90
ORIENT	Orientation Test	24
MAP	Map Test	20
TARGET1	Target Tracking Test 1	18
TARGET2	Target Tracking Test 2	18
IDENT.D	Target Identification, Time	36
IDENT.H	Target Identification, Hits	30
MEM.DIS	Memory Search Test, Time	
MEM.HIT	Memory Search Test, Hits	
ECAT subtests		
ID	Integrating Details	40
SM	Sequential Memory	35
AO	Assembling Objects	32
FR	Figural Reasoning	35
SO	Spatial Orientation	24
T1	One-Hand Tracking	18
T2	Two-Hand Tracking	18
TI	Target Identification	36

Table 11
Factor Scores Description

1. Infantryman (n = 491)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	48.00	4.92	34.89	60.06	0.927
Gc	14.32	4.54	1.15	29.17	0.900
Gv/Mech	14.86	4.30	0.15	26.94	0.865
Speed	15.40	3.51	4.39	24.85	0.840
Math	-5.43	3.64	-18.15	6.69	0.776
Psymotor	11.60	1.09	8.89	15.06	0.634
G	54.10	5.42	40.19	66.32	
Hands-On	54.15	5.68	36.50	69.98	

2. Cannon Crewman (n = 464)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	46.19	4.50	32.99	58.33	0.927
Gc	12.80	5.16	-8.74	27.93	0.900
Gv/Mech	12.99	4.81	-1.71	26.37	0.865
Speed	15.31	3.62	2.47	26.63	0.840
Math	-5.24	3.44	-13.71	7.93	0.776
Psymotor	11.80	1.22	8.74	15.98	0.634
G	50.90	5.68	37.04	67.84	
Hands-On	48.01	8.72	20.62	73.86	

3. Tank Crewman (n = 394)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	47.65	4.72	34.74	58.39	0.927
Gc	14.25	4.93	-2.76	25.56	0.900
Gv/Mech	15.34	4.27	0.41	25.95	0.865
Speed	14.88	3.64	5.18	24.37	0.840
Math	-5.17	3.44	-15.74	5.65	0.776
Psymotor	11.62	1.08	9.05	14.90	0.634
G	53.84	5.60	41.00	67.53	
Hands-On	59.87	5.65	39.88	72.10	

Table 11
Factor Scores Description (continued)

4. Radio Operator (n = 289)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	48.25	4.52	37.50	59.40	0.927
Gc	13.80	4.53	0.85	26.33	0.900
Gv/Mech	13.42	5.14	-3.14	26.75	0.865
Speed	17.33	3.34	7.65	24.80	0.840
Math	-5.67	3.58	-15.67	4.23	0.776
Psymotor	11.76	1.07	9.00	14.87	0.634
G	53.62	5.15	41.49	65.38	
Hands-On	54.19	6.03	33.60	68.37	

5. Vehicle Mechanic (n = 478)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	46.76	4.54	35.51	57.97	0.927
Gc	13.34	4.36	-0.27	24.72	0.900
Gv/Mech	16.60	4.65	2.45	27.80	0.865
Speed	15.46	3.51	5.87	25.93	0.840
Math	-5.53	3.49	-17.53	4.27	0.776
Psymotor	11.93	1.13	8.64	15.83	0.634
G	52.96	5.25	38.81	65.62	
Hands-On	65.14	3.70	46.20	72.75	

6. Motor Transport (n = 507)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	45.51	4.66	33.03	58.21	0.927
Gc	12.68	4.67	-1.68	24.57	0.900
Gv/Mech	15.18	4.96	-3.05	26.72	0.865
Speed	15.42	3.64	5.38	25.94	0.840
Math	-5.84	3.30	-16.58	4.32	0.776
Psymotor	11.85	1.18	8.78	15.58	0.634
G	50.86	5.06	39.39	65.14	
Hands-On	55.10	5.81	33.07	68.65	

Table 11
Factor Scores Description (continued)

7. Administrative (n = 427)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	47.64	4.67	33.92	59.85	0.927
Gc	13.20	4.18	-2.38	23.84	0.900
Gv/Mech	9.14	4.93	-4.98	25.06	0.865
Speed	18.09	3.08	7.67	25.21	0.840
Math	-5.66	3.72	-17.98	6.39	0.776
Psymotor	12.27	1.16	8.95	15.91	0.634
G	51.04	5.32	38.22	65.38	
Hands-On	45.63	6.32	25.55	63.64	

8. Medical (n = 392)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	49.15	4.17	35.99	59.49	0.927
Gc	15.25	4.09	-1.38	28.62	0.900
Gv/Mech	11.98	5.14	-1.09	24.42	0.865
Speed	15.14	3.92	3.80	26.49	0.840
Math	-5.44	3.52	-14.63	4.69	0.776
Psymotor	11.95	1.17	8.97	15.22	0.634
G	54.28	4.40	44.55	65.34	
Hands-On	55.77	5.38	37.49	69.56	

9. Military Police (n = 597)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	49.32	3.68	37.81	58.53	0.927
Gc	15.60	3.71	2.98	26.44	0.900
Gv/Mech	14.95	4.42	-0.09	27.07	0.865
Speed	15.17	3.46	3.37	25.08	0.840
Math	-5.03	3.24	-14.41	6.62	0.776
Psymotor	11.55	1.02	8.70	15.19	0.634
G	56.02	3.67	46.38	65.27	
Hands-On	54.12	4.71	37.99	64.87	

Table 11
Factor Scores Description (continued)

10. Helicopter Mechanic (n = 439)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	27.81	4.44	14.35	37.28	0.946
Gc	25.83	4.04	7.81	34.73	0.902
Gv/Mech	24.67	4.14	13.67	34.36	0.861
Speed	30.49	3.70	8.01	41.93	0.859
Math	5.42	2.94	-4.15	13.27	0.742
Psymotor	-57.71	6.54	-79.35	-42.60	0.916
G	50.78	3.90	40.22	61.45	
Hands-On	76.92	7.98	45.00	94.00	

11. Automotive Mechanic (n = 694)

Variable	Mean	Std Dev	Minimum	Maximum	Determinacy
Gf	29.32	4.40	18.51	41.57	0.935
Gc	21.67	4.21	-0.08	33.49	0.887
Gv/Mech	20.42	4.03	6.12	30.77	0.858
Speed	32.65	3.70	20.50	42.17	0.853
Math	4.74	3.09	-5.14	13.99	0.747
Psymotor	-58.84	6.88	-83.29	-41.49	0.883
G	48.58	4.22	37.62	61.79	
Hands-On	77.82	7.61	43.00	93.00	

Table 12
Estimates for the Standardized Prediction Equation and Adjusted R Squares

Job	N	Gf	Gc	Gv/Mech	Speed	Math	Psymotor	Adjusted R^2
Project A								
1. Infantryman	491	0.258[a]	0.081	0.218	0.023	0.079	-0.075	0.178
		5.922[b]	1.854	5.087	0.512	1.764	-1.679	
2. Cannon Crewman	464	0.140	0.034	0.113	0.001	0.021	-0.027	0.050
		2.780	0.720	2.367	0.020	0.433	-0.566	
3. Tank Crewman	394	0.242	0.021	0.223	-0.075	0.080	-0.028	0.190
		4.853	0.431	4.645	-1.488	1.596	-0.562	
4. Radio Operator	289	0.338	0.213	0.205	0.158	0.090	0.057	0.229
		5.034	3.271	3.339	2.112	1.518	0.983	
5. Vehicle Mechanic	478	0.117	0.057	0.236	-0.093	0.046	-0.110	0.161
		2.581	1.256	5.264	-1.983	0.990	-2.304	
6. Motor Transport	507	0.318	0.069	0.283	-0.008	0.002	-0.024	0.247
		7.524	1.533	6.262	-0.181	0.051	-0.562	
7. Administrative	427	0.469	-0.026	0.055	0.054	0.041	0.053	0.348
		9.677	-0.577	1.192	1.097	0.905	1.174	
8. Medical	392	0.362	0.131	0.223	0.086	0.045	0.004	0.270
		7.033	2.524	4.666	1.724	0.880	0.084	
9. Military Police	597	0.323	0.063	0.258	-0.014	0.130	0.000	0.290
		7.779	1.470	6.716	-0.357	3.327	0.003	
Marine Corps								
10. Helicopter Mechanic	439	0.404	-0.011	0.295	0.036	0.030	-0.024	0.519
		7.983	-0.226	6.056	0.809	0.659	-0.530	
11. Automotive Mechanic	694	0.355	0.135	0.343	0.006	-0.005	-0.024	0.569
		9.667	3.763	10.138	0.171	-0.136	-0.711	

[a]Standardized regression coefficient.
[b]*t* ratio.

Table 13
Estimates for the Standardized Prediction Equation (Top 50 %)

Job	N	Gf	Gc	Gv/Mech	Speed	Math	Psymotor
Project A							
1. Infantryman	246	0.228[a]	0.123	0.177	0.026	-0.028	0.040
		3.252[b]	1.675	2.507	0.403	-0.407	0.571
2. Cannon Crewman	232	0.004	-0.088	0.061	-0.052	0.028	-0.047
		0.064	-1.246	0.875	-0.737	0.394	-0.662
3. Tank Crewman	197	0.036	0.085	0.248	-0.099	0.120	-0.088
		0.450	1.163	3.268	-1.354	1.591	-1.186
4. Radio Operator	145	0.200	0.220	0.172	0.085	0.121	0.089
		1.950	2.267	1.870	0.905	1.387	1.003
5. Vehicle Mechanic	239	0.086	-0.034	0.116	-0.162	0.065	-0.088
		1.246	-0.486	1.667	-2.359	0.933	-1.223
6. Motor Transport	254	0.261	0.144	0.324	0.006	-0.055	-0.051
		3.982	2.082	4.620	0.087	-0.869	-0.781
7. Administrative	214	0.452	0.029	0.144	0.048	0.023	0.041
		6.537	0.422	2.199	0.734	0.351	0.635
8. Medical	196	0.288	0.064	0.171	0.106	0.124	-0.007
		3.280	0.822	2.077	1.471	1.595	-0.084
9. Military Police	299	0.203	0.011	0.142	-0.074	0.241	-0.026
		3.007	0.175	2.278	-1.290	3.983	-0.437
Marine Corps							
10. Helicopter Mechanic	220	0.331	0.034	0.315	0.012	0.017	0.062
		4.397	0.456	4.457	0.185	0.256	0.920
11. Automotive Mechanic	347	0.268	0.156	0.354	-0.003	0.014	-0.002
		4.557	2.763	6.885	-0.062	0.275	-0.035

[a]Standardized regression coefficient.
[b]*t* ratio.

Table 14
Spearman Rank-Order Correlations
(Predicted Y with Hands-On Job Performance)

Job	N	$\hat{g}+5\hat{s}$	$\hat{g}+\hat{s}^*$	$\hat{g}_1$
Project A				
1. Infantryman	491	0.351[a]	0.338	0.331
		8.277[b]	7.946	7.754
2. Cannon Crewman	464	0.222	0.216	0.218
		4.901	4.743	4.803
3. Tank Crewman	394	0.375	0.355	0.337
		8.014	7.522	7.085
4. Radio Operator	289	0.402	0.390	0.391
		7.430	7.177	7.206
5. Vehicle Mechanic	478	0.322	0.308	0.267
		7.410	7.059	6.043
6. Motor Transport	507	0.414	0.409	0.365
		10.220	10.070	8.822
7. Administrative	427	0.439	0.430	0.374
		10.079	9.830	8.320
8. Medical	392	0.377	0.368	0.369
		8.044	7.815	7.842
9. Military Police	597	0.394	0.391	0.367
		10.464	10.364	9.613
Marine Corps				
10. Helicopter Mechanic	439	0.400	0.400	0.293
		9.113	9.111	6.415
11. Automotive Mechanic	694	0.488	0.487	0.395
		14.724	14.651	11.327

[a]Spearman rank-order correlation.
[b]t ratio.

Table 15
Job Classifications

Job (Top 50%)	N	Selection Method	Failure Reject	Failure Accept	Success Reject	Success Accept	Sensitivity	Specificity	Proportion of Correct Decisions
Project A									
1. Infantryman	491	$\hat{g}+5\hat{s}$	148	97	97	149	60.57%	60.41%	60.49%
		$\hat{g}+\hat{s}^*$	149	96	96	150	60.98%	60.82%	60.90%
		$\hat{g}_1$	149	96	96	150	60.98%	60.82%	60.90%
2. Cannon Crewman	464	$\hat{g}+5\hat{s}$	125	107	107	125	53.88%	53.88%	53.88%
		$\hat{g}$	141	91	91	141	60.78%	60.78%	60.78%
		$\hat{g}_1$	140	92	92	140	60.34%	60.34%	60.34%
3. Tank Crewman	394	$\hat{g}+5\hat{s}$	123	74	74	123	62.44%	62.44%	62.44%
		$\hat{g}+\hat{s}^*$	120	77	77	120	60.91%	60.91%	60.91%
		$\hat{g}_1$	125	72	72	125	63.45%	63.45%	63.45%
4. Radio Operator	289	$\hat{g}+5\hat{s}$	95	49	49	96	66.21%	65.97%	66.09%
		$\hat{g}+\hat{s}^*$	88	56	56	89	61.38%	61.11%	61.25%
		$\hat{g}_1$	96	48	48	97	66.90%	66.67%	66.78%
5. Vehicle Mechanic	478	$\hat{g}+5\hat{s}$	155	84	84	155	64.85%	64.85%	64.85%
		$\hat{g}+\hat{s}^*$	144	95	95	144	60.25%	60.25%	60.25%
		$\hat{g}_1$	148	91	91	148	61.92%	61.92%	61.92%
6. Motor Transport	507	$\hat{g}+5\hat{s}$	160	93	93	161	63.39%	63.24%	63.31%
		$\hat{g}+\hat{s}^*$	154	99	99	155	61.02%	60.87%	60.95%
		$\hat{g}_1$	158	95	95	159	62.60%	62.45%	62.52%
7. Administrative	427	$\hat{g}+5\hat{s}$	136	77	77	137	64.02%	63.85%	63.93%
		$\hat{g}+\hat{s}^*$	135	78	78	136	63.55%	63.38%	63.47%
		$\hat{g}_1$	128	85	85	129	60.28%	60.09%	60.19%
8. Medical	392	$\hat{g}+5\hat{s}$	123	73	73	123	62.76%	62.76%	62.76%
		$\hat{g}+\hat{s}^*$	121	75	75	121	61.73%	61.73%	61.73%
		$\hat{g}_1$	122	74	74	122	62.24%	62.24%	62.24%
9. Military Police	597	$\hat{g}+5\hat{s}$	191	107	107	192	64.21%	64.09%	64.15%
		$\hat{g}+\hat{s}^*$	186	112	112	187	62.54%	62.42%	62.48%
		$\hat{g}_1$	186	112	112	187	62.54%	62.42%	62.48%
Marine Corps									
10. Helicopter Mechanic	439	$\hat{g}+5\hat{s}$	140	79	79	141	64.09%	63.93%	64.01%
		$\hat{g}+\hat{s}^*$	140	79	79	141	64.09%	63.93%	64.01%
		$\hat{g}_1$	131	88	88	132	60.00%	59.82%	59.91%
11. Automotive Mechani	694	$\hat{g}+5\hat{s}$	240	107	107	240	69.16%	69.16%	69.16%
		$\hat{g}+\hat{s}^*$	240	107	107	240	69.16%	69.16%	69.16%
		$\hat{g}_1$	225	122	122	225	64.84%	64.84%	64.84%

Note. Regression based on selected sample from single-factor model.

Figure 1
Observed Data Pattern with Missing Data on Criterion

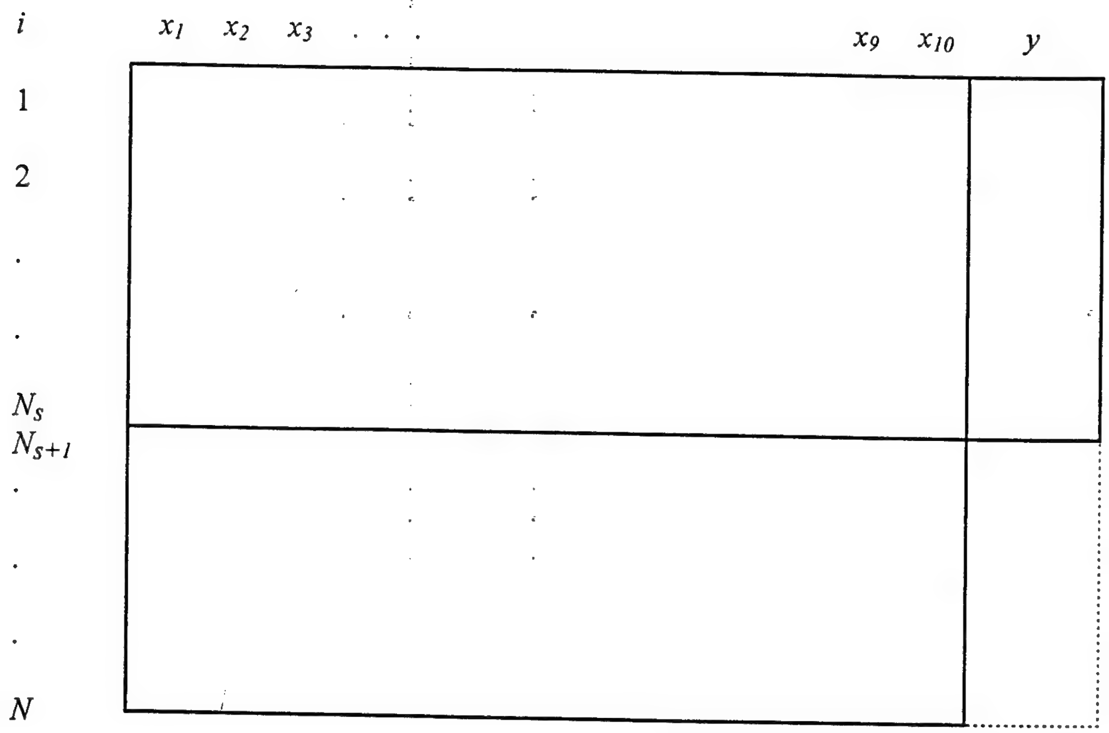


Figure 2
General Form of the Decision Table

Future Performance	Failure	TN	FP
	Success	FN	TP
		Reject	Accept

Selection Decision

Sensitivity = TP / (FN + TP)

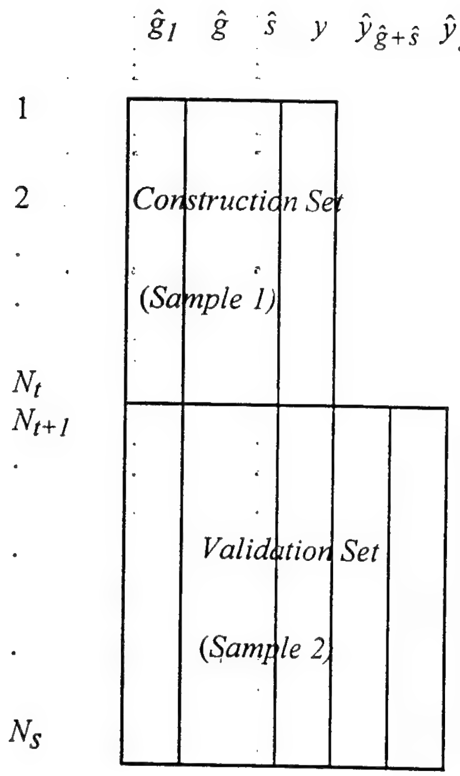
Specificity = TN / (TN + FP)

Proportion of Correct Decisions

= (TP + TN) / N

= (TP + TN) / (FN + TP + TN + FP)

Figure 3
Data Pattern for Cross-Validation in the Selected Sample



Estimates of Regression Slope

$$y = \hat{\alpha} + \hat{\beta}_g \times \hat{g} + \hat{\beta}_s \times \hat{s}$$

$$y = \hat{\alpha} + \hat{\beta}_{g_1} \times \hat{g}_1$$

Estimates of Criterion

$$\hat{y}_{\hat{g}+\hat{s}} = \hat{\alpha} + \hat{\beta}_g \times \hat{g} + \hat{\beta}_s \times \hat{s}$$

$$\hat{y}_{\hat{g}_1} = \hat{\alpha} + \hat{\beta}_{g_1} \times \hat{g}_1$$

$$PRESS = \sqrt{\frac{\Sigma(\hat{y} - y)^2}{N_t}}$$

Note: N_S represents the sample size for the selected subjects.
$N_t = N_S/2$.

Figure 4
Single-Factor Model

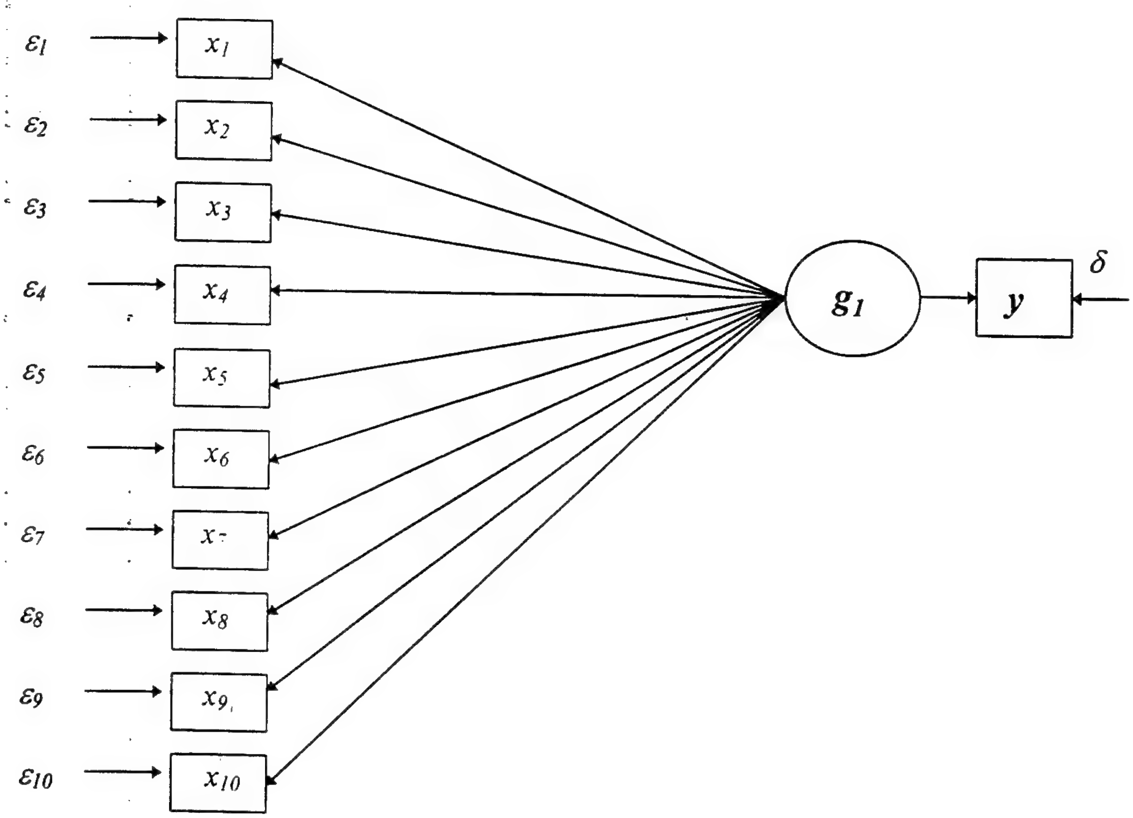


Figure 5
Hierarchical Factor Model

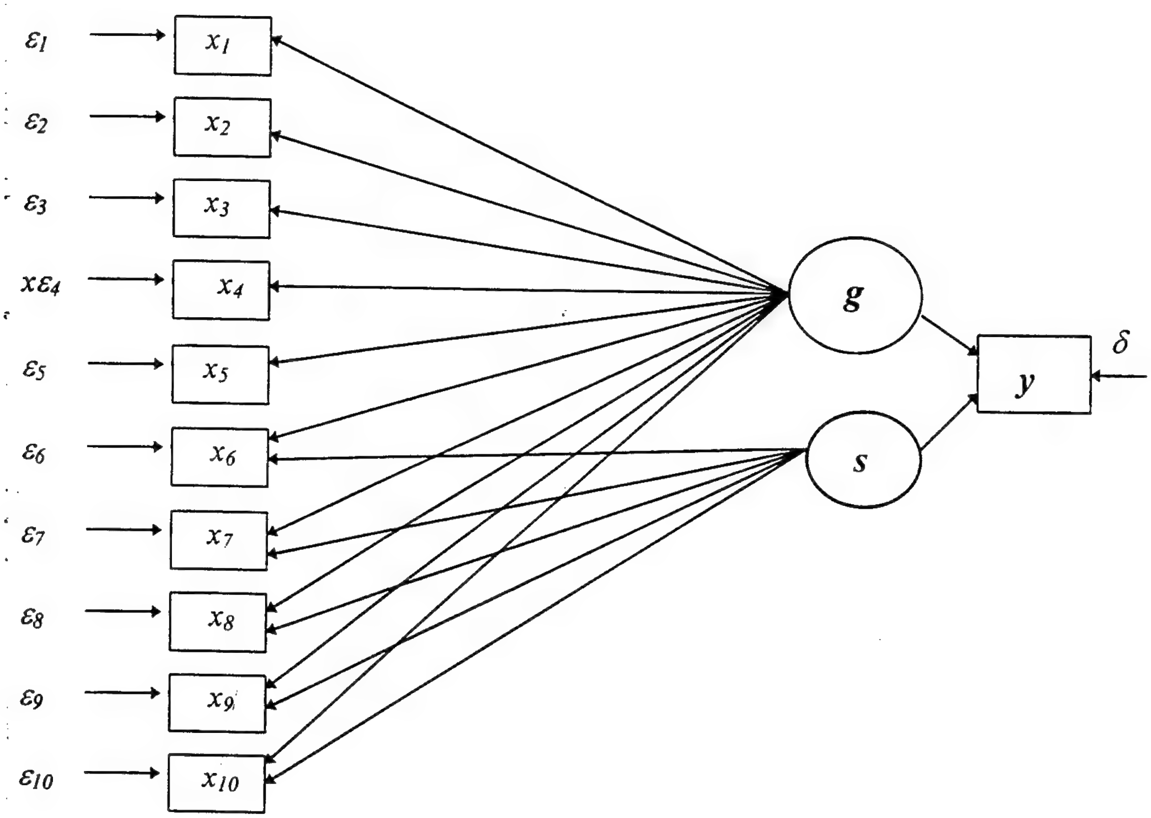


Figure 6
Data Pattern for the Artificial Data

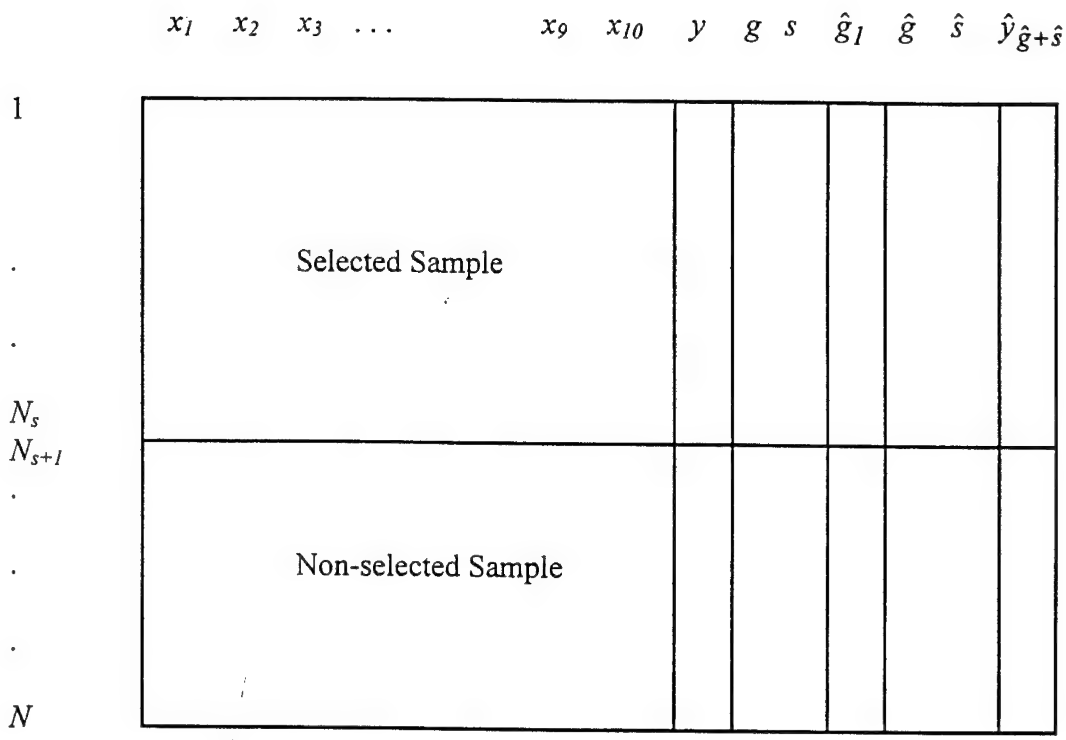
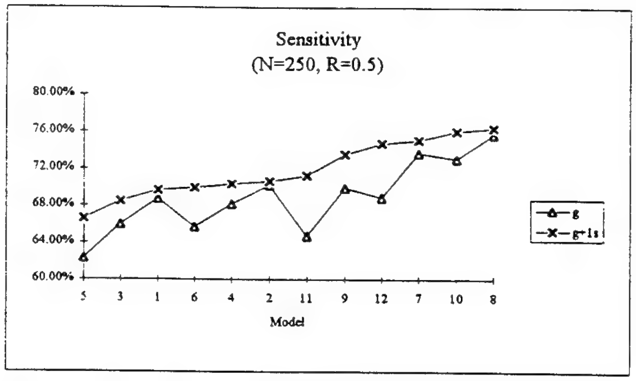
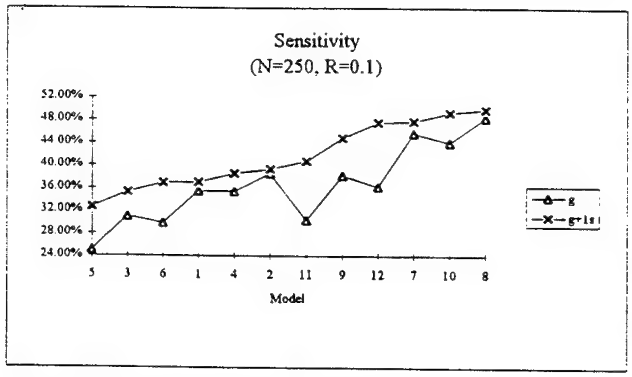
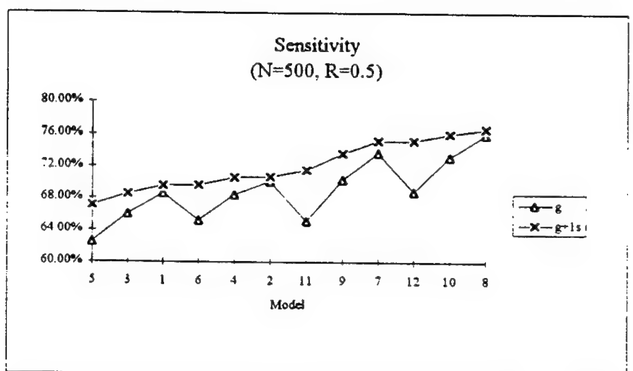


Figure 7
Plots of Classification for $g + 1\ s$







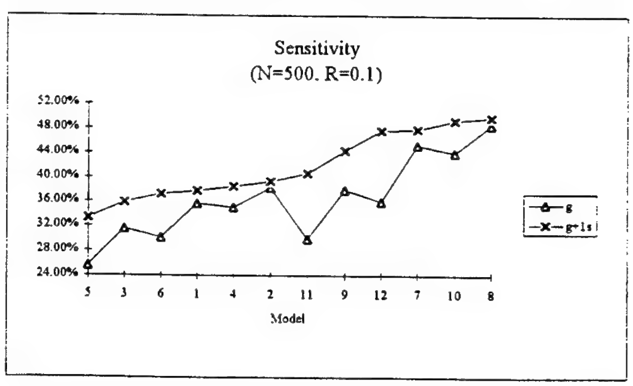


Figure 7
Plots of Classification for $g + 1\,s$

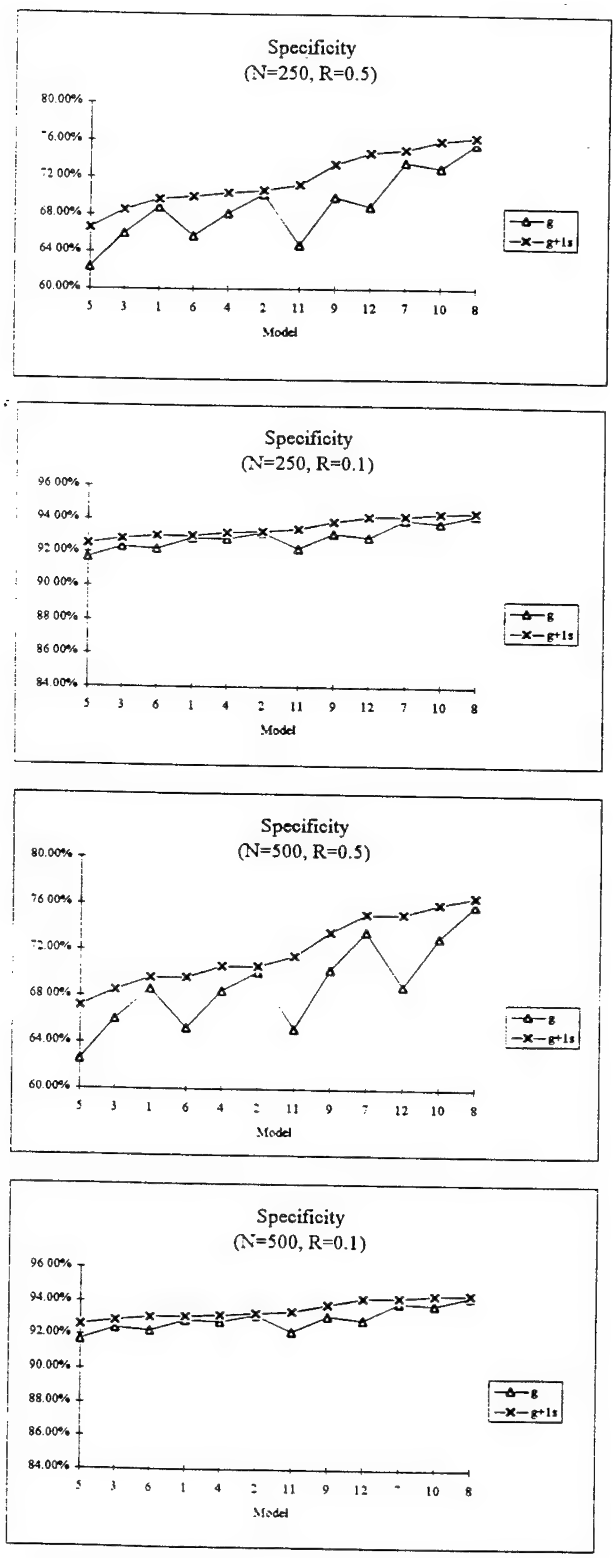
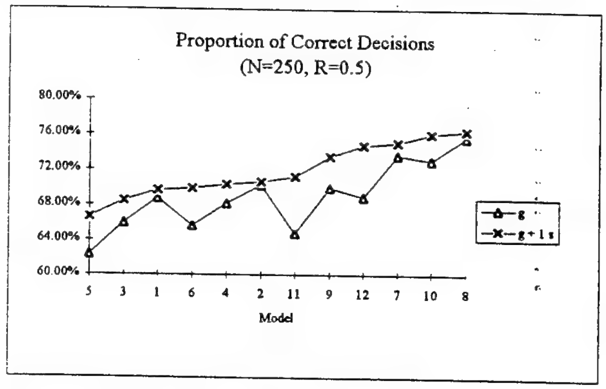
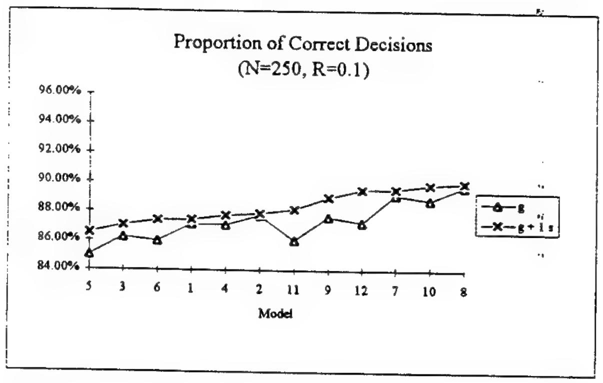
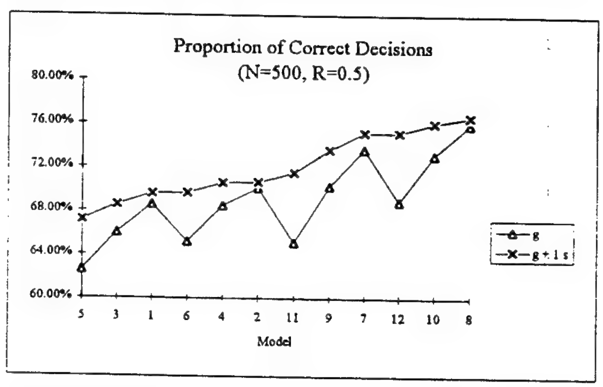


Figure 7
Plots of Classification for $g + 1s$







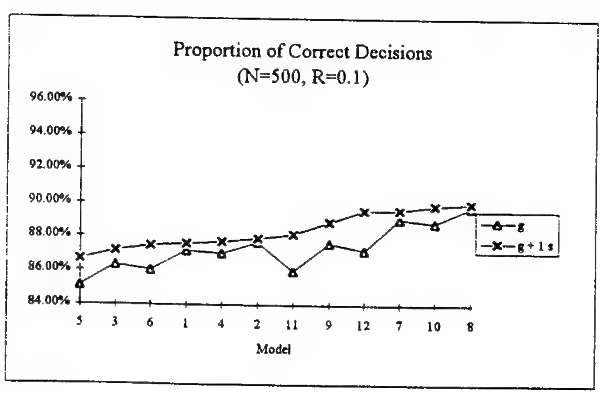


Figure 8
Plots of Classification for $g + 2s$

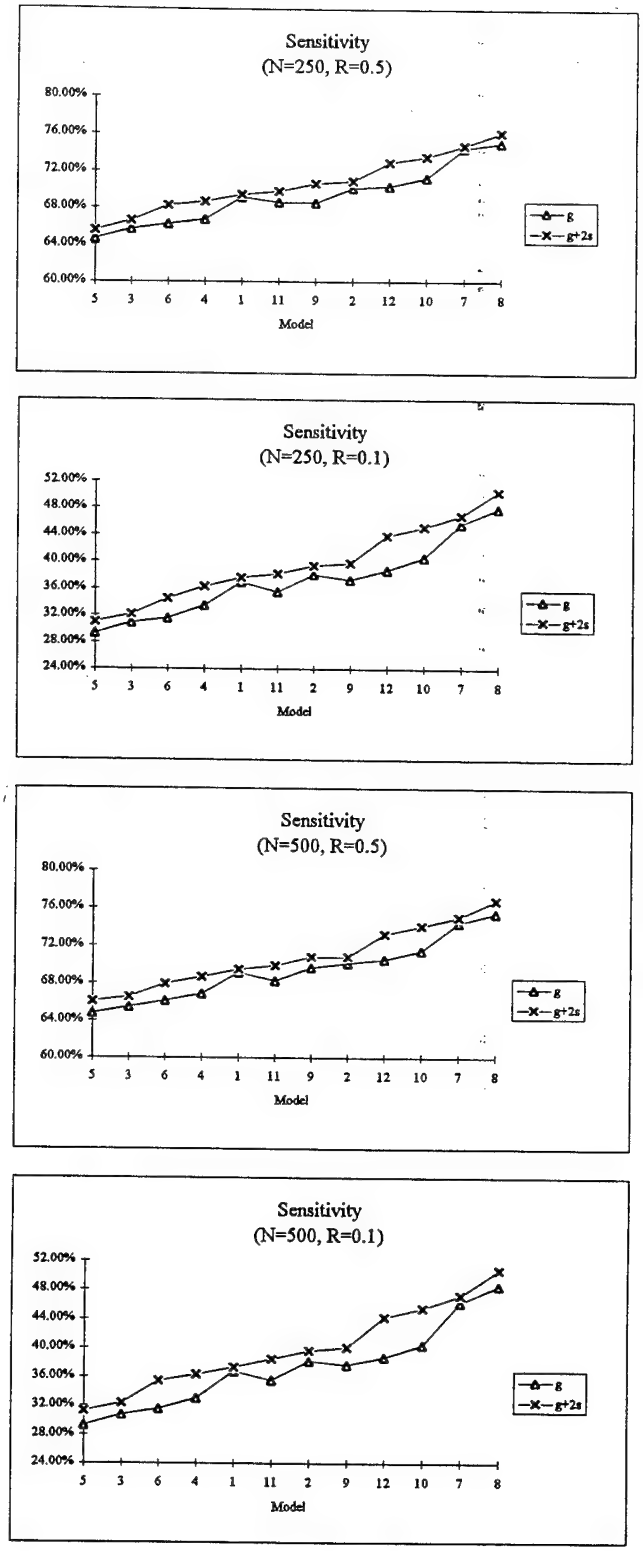
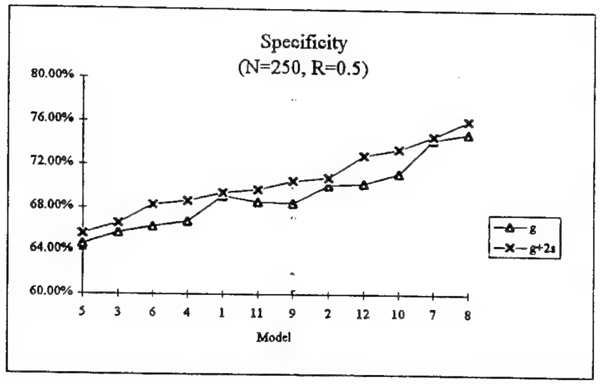
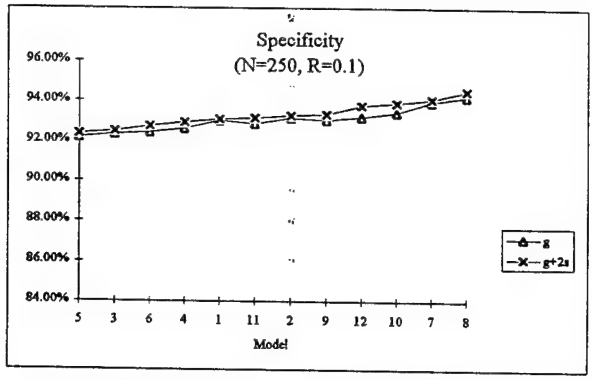
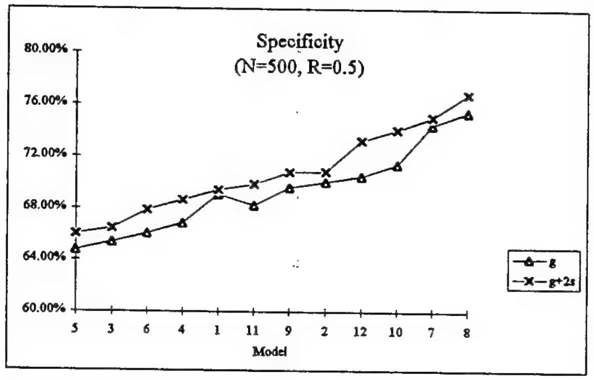


Figure 8
Plots of Classification for $g + 2s$







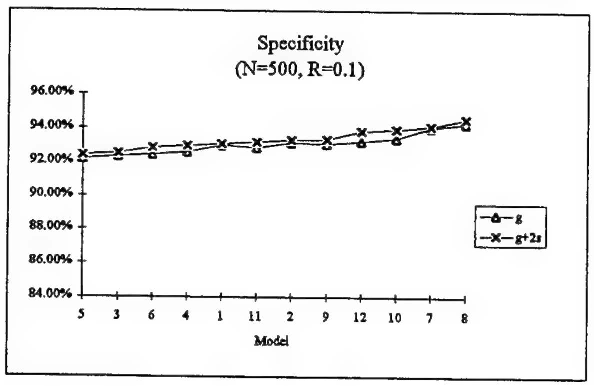
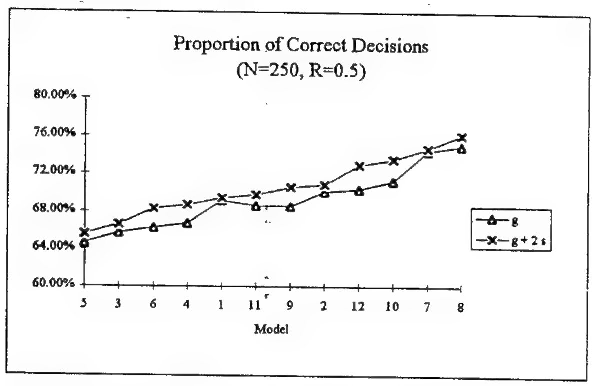
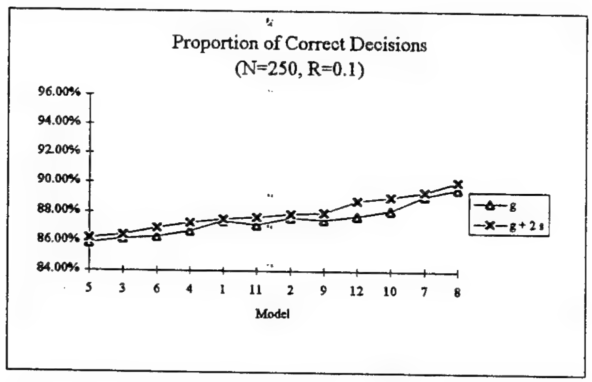
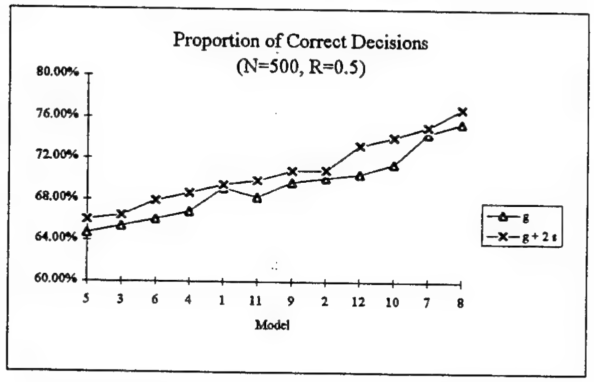


Figure 8
Plots of Classification for $g + 2s$







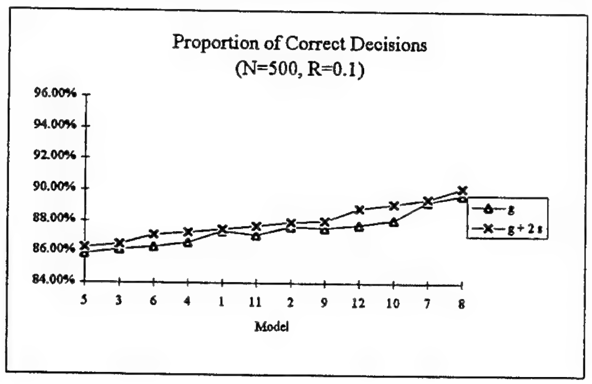
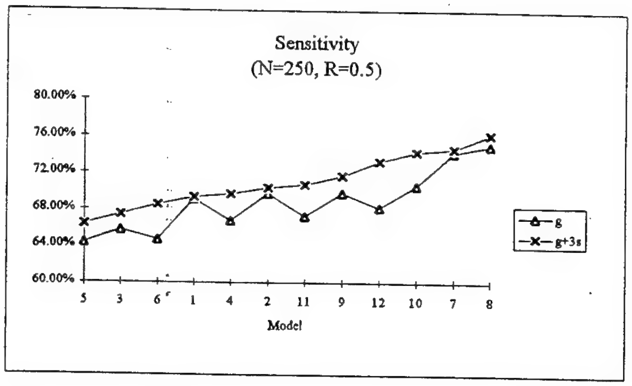
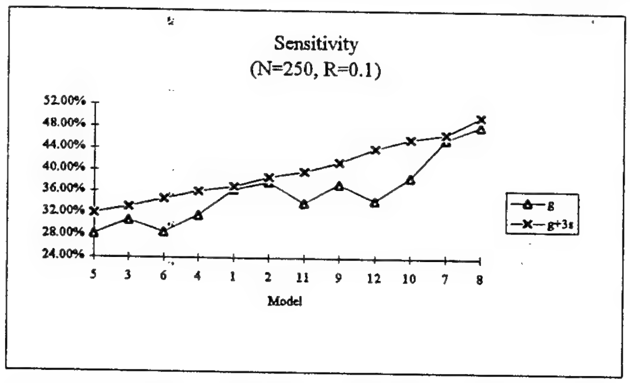
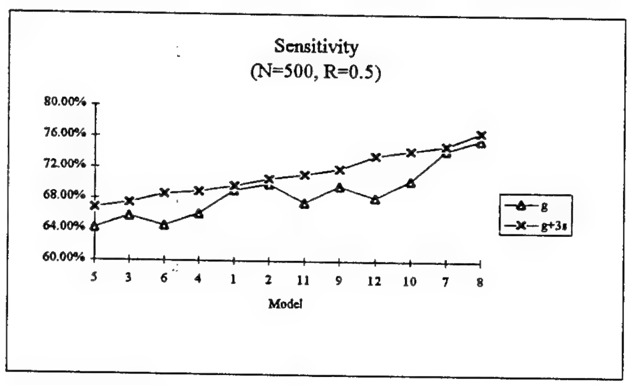


Figure 9
Plots of Classification for $g + 3s$







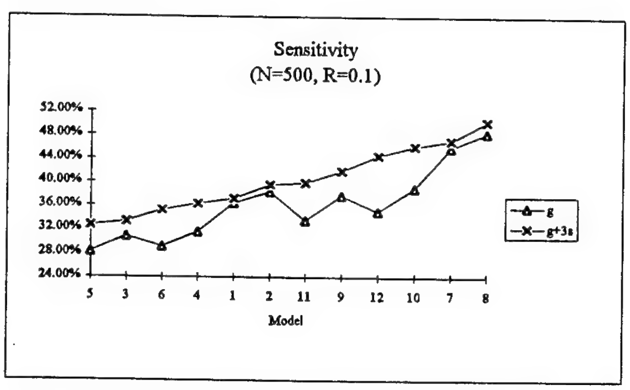
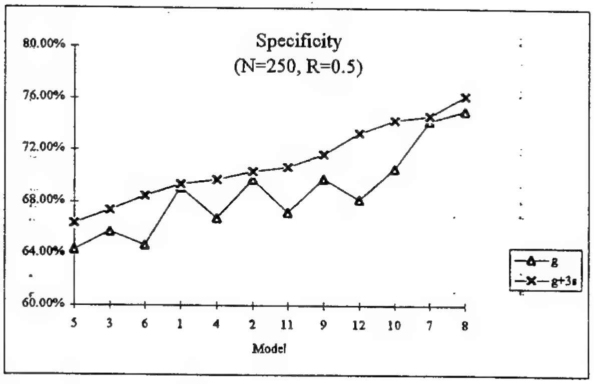
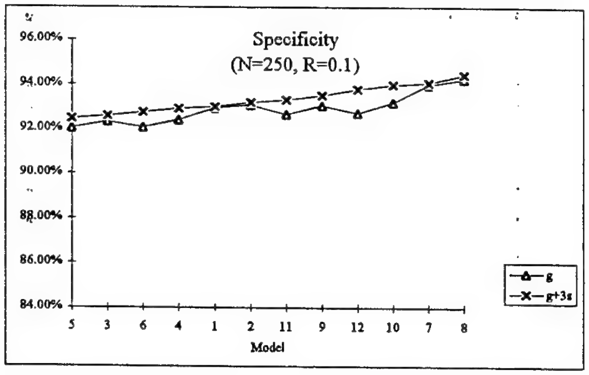
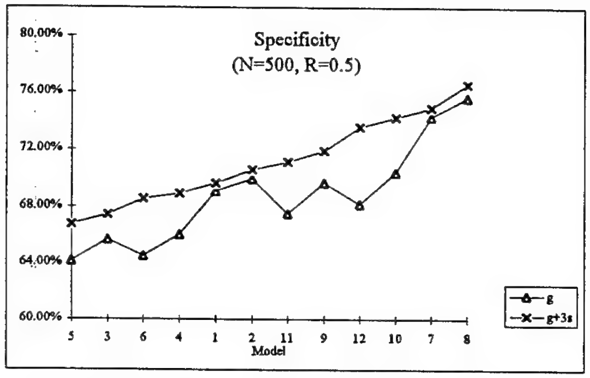


Figure 9
Plots of Classification for $g + 3s$







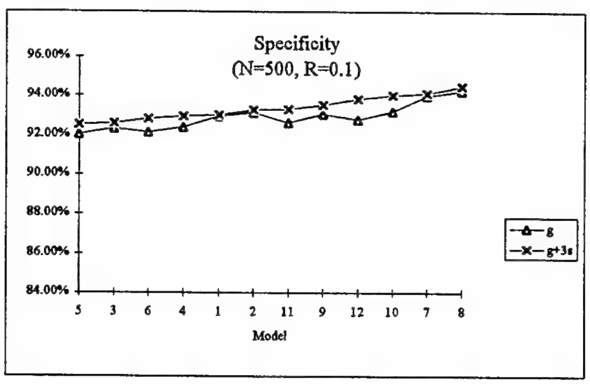
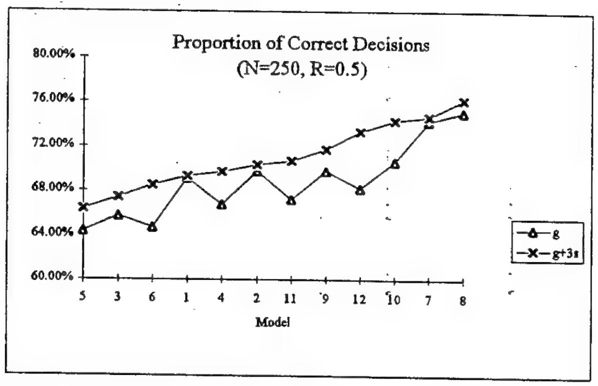
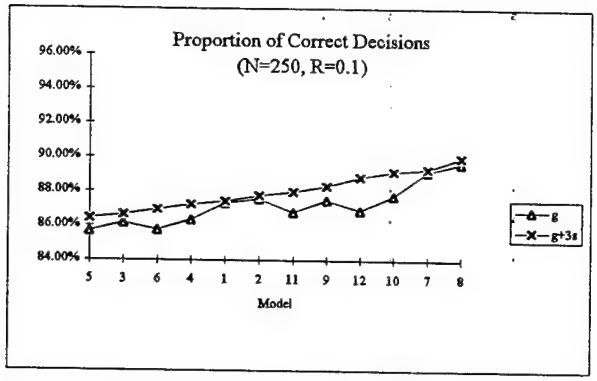
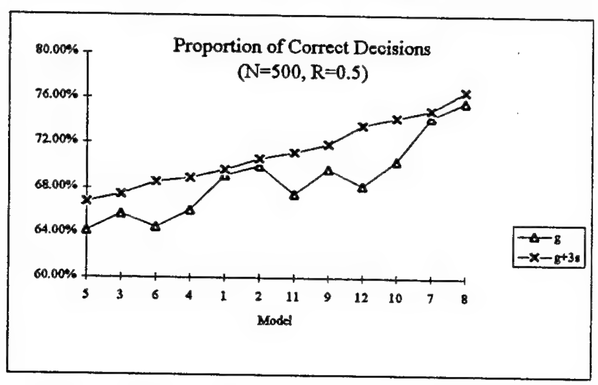


Figure 9
Plots of Classification for $g + 3\ s$







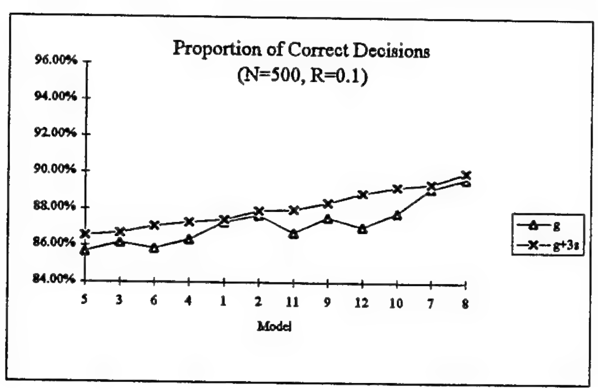
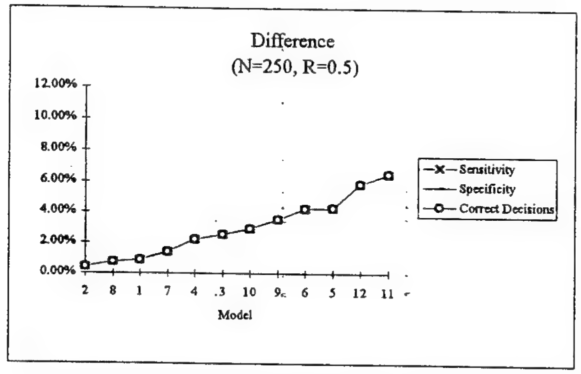
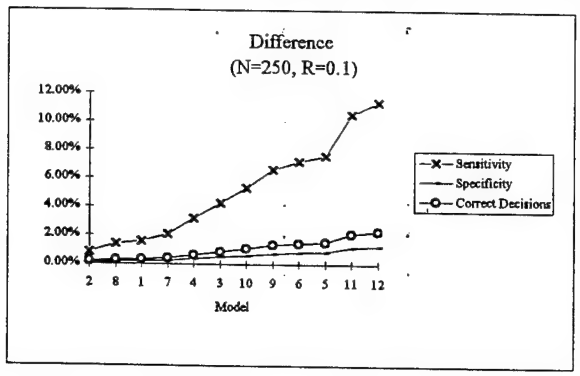
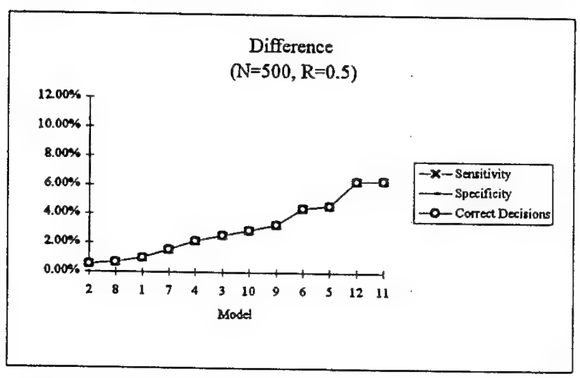


Figure 10
Plots of Difference for $g + 1s$ Classification







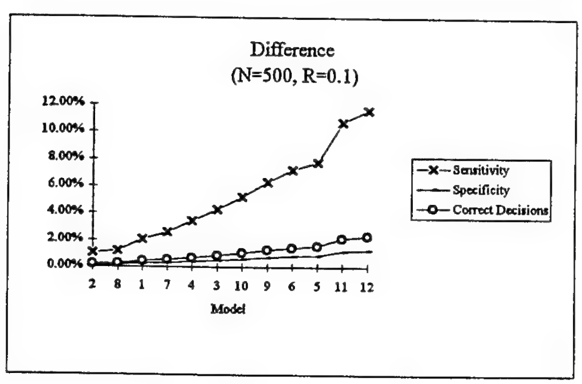
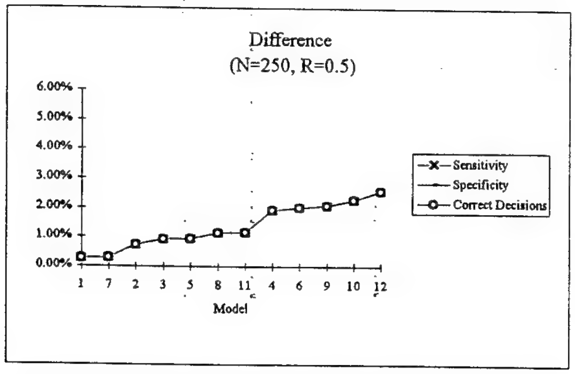
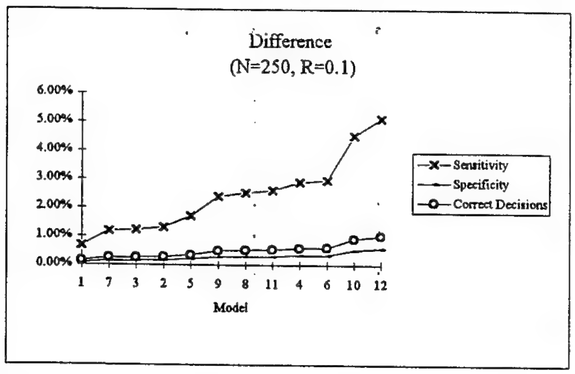
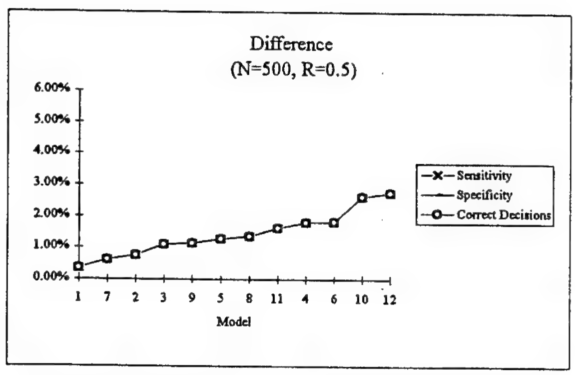


Figure 11
Plots of Difference for $g + 2s$ Classification







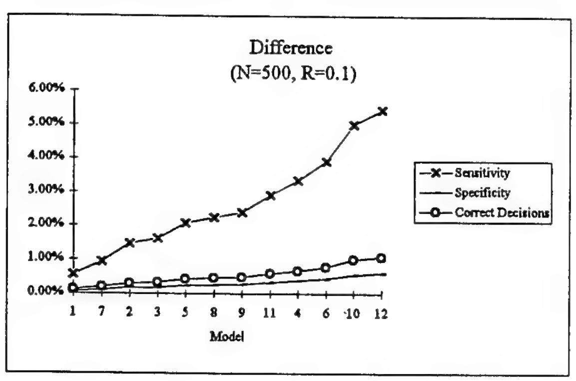


Figure 12
Plots of Difference for $g + 3s$ Classification

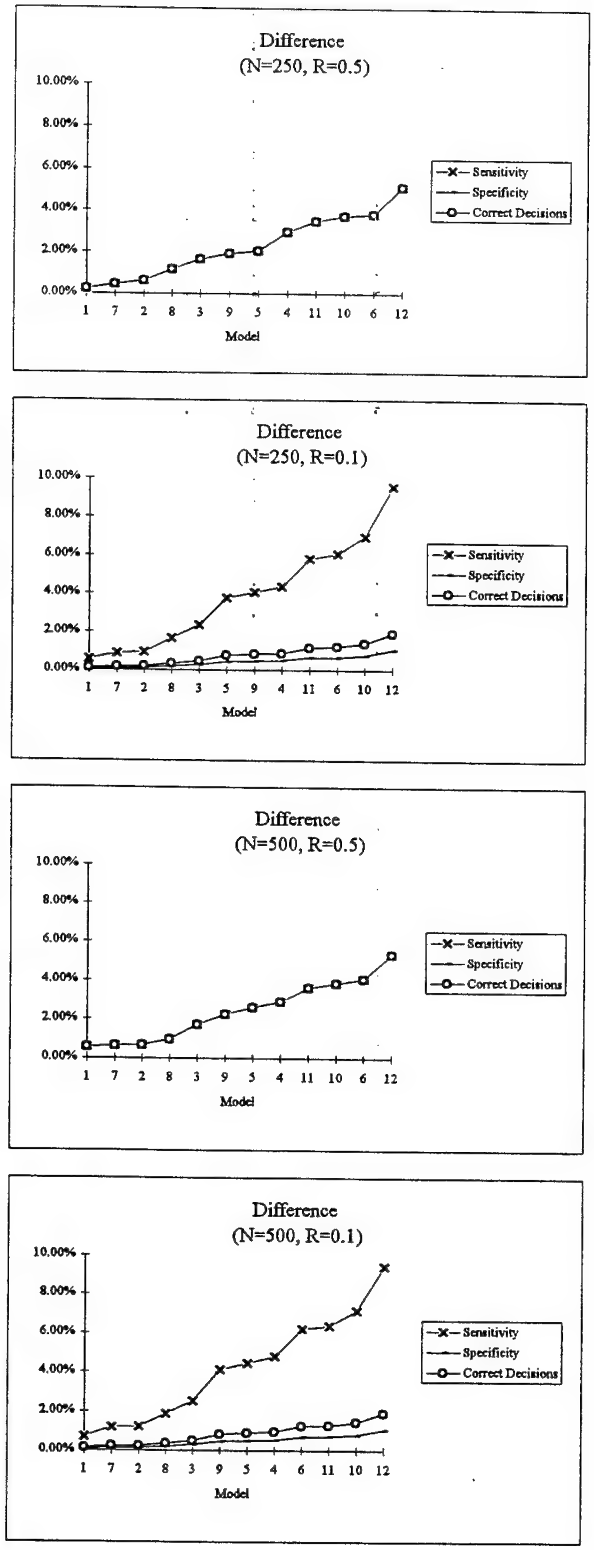


Figure 13
Plots of Spearman Rank-Order Correlation for $g + 1s$

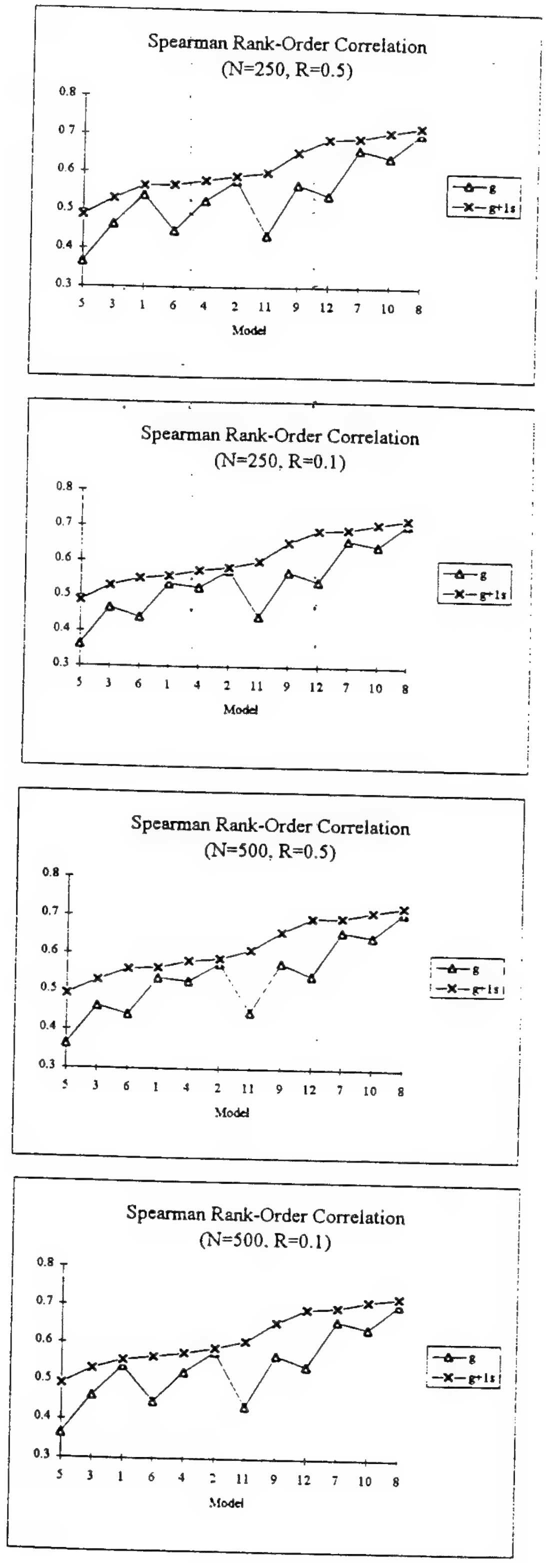
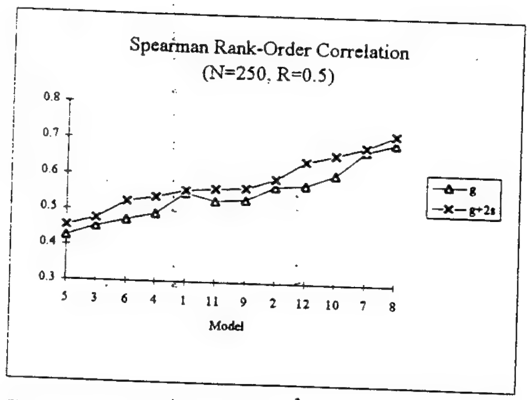
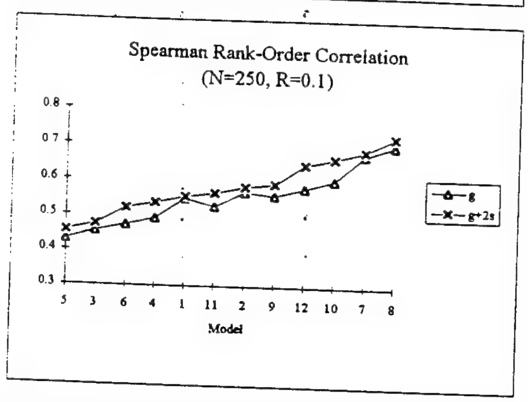
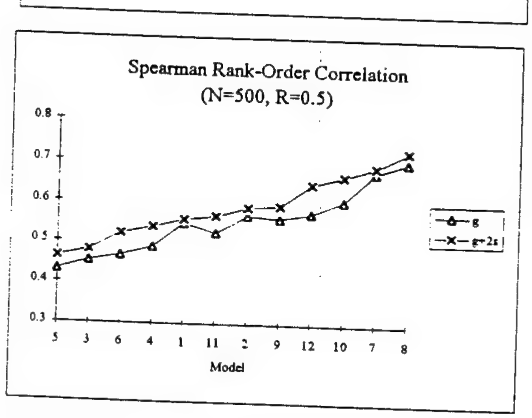


Figure 14
Plots of Spearman Rank-Order Correlation for $g + 2s$







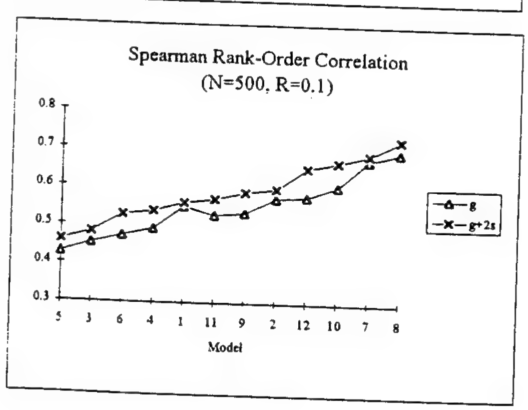
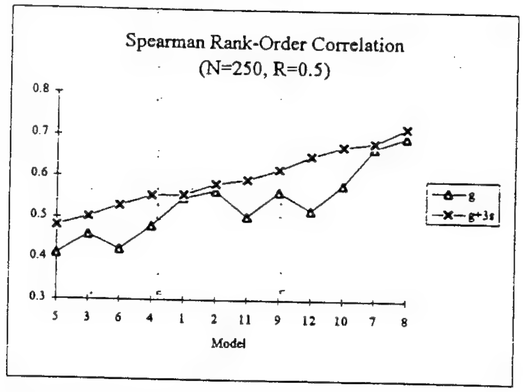
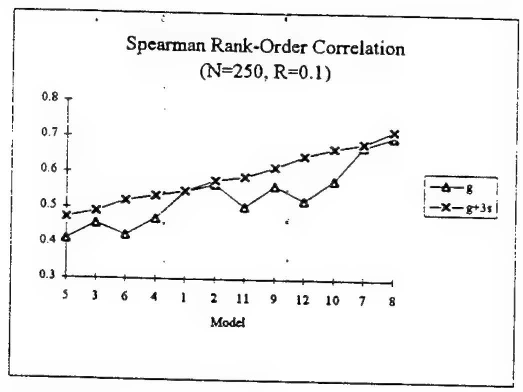
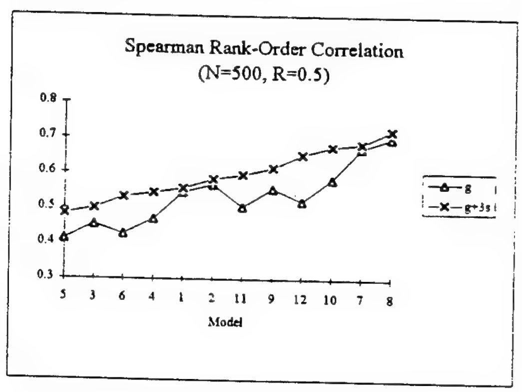


Figure 15

Plots of Spearman Rank-Order Correlation for $g + 3s$







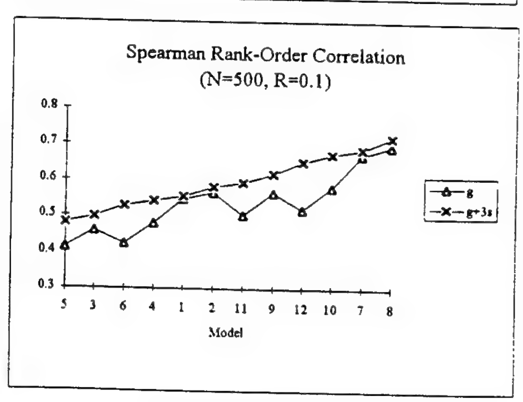


Figure 16
Plots of Difference for Cross-Validation of $g + 1$ s

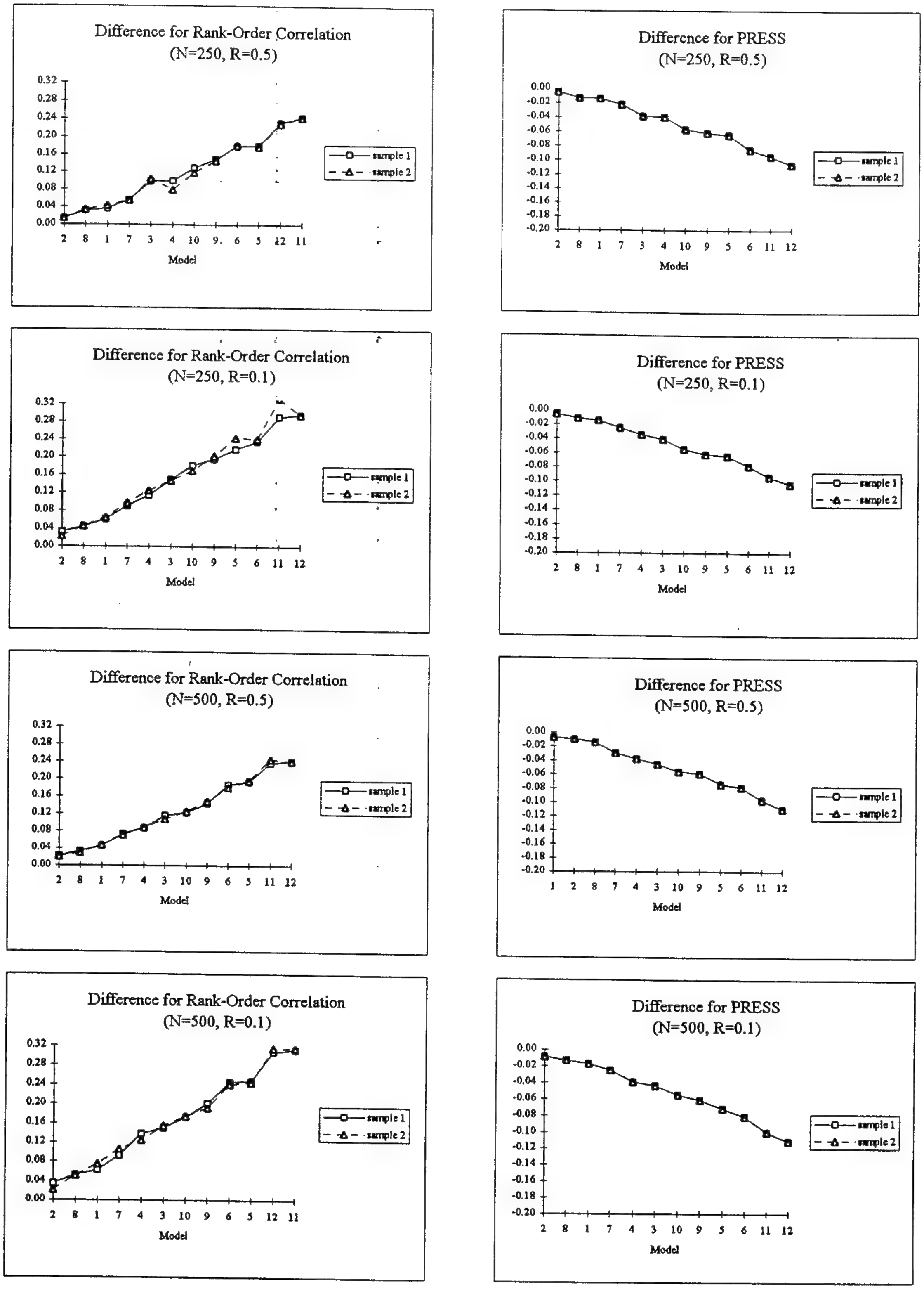


Figure 17
Plots of Difference for Cross-Validation of $g + 2s$

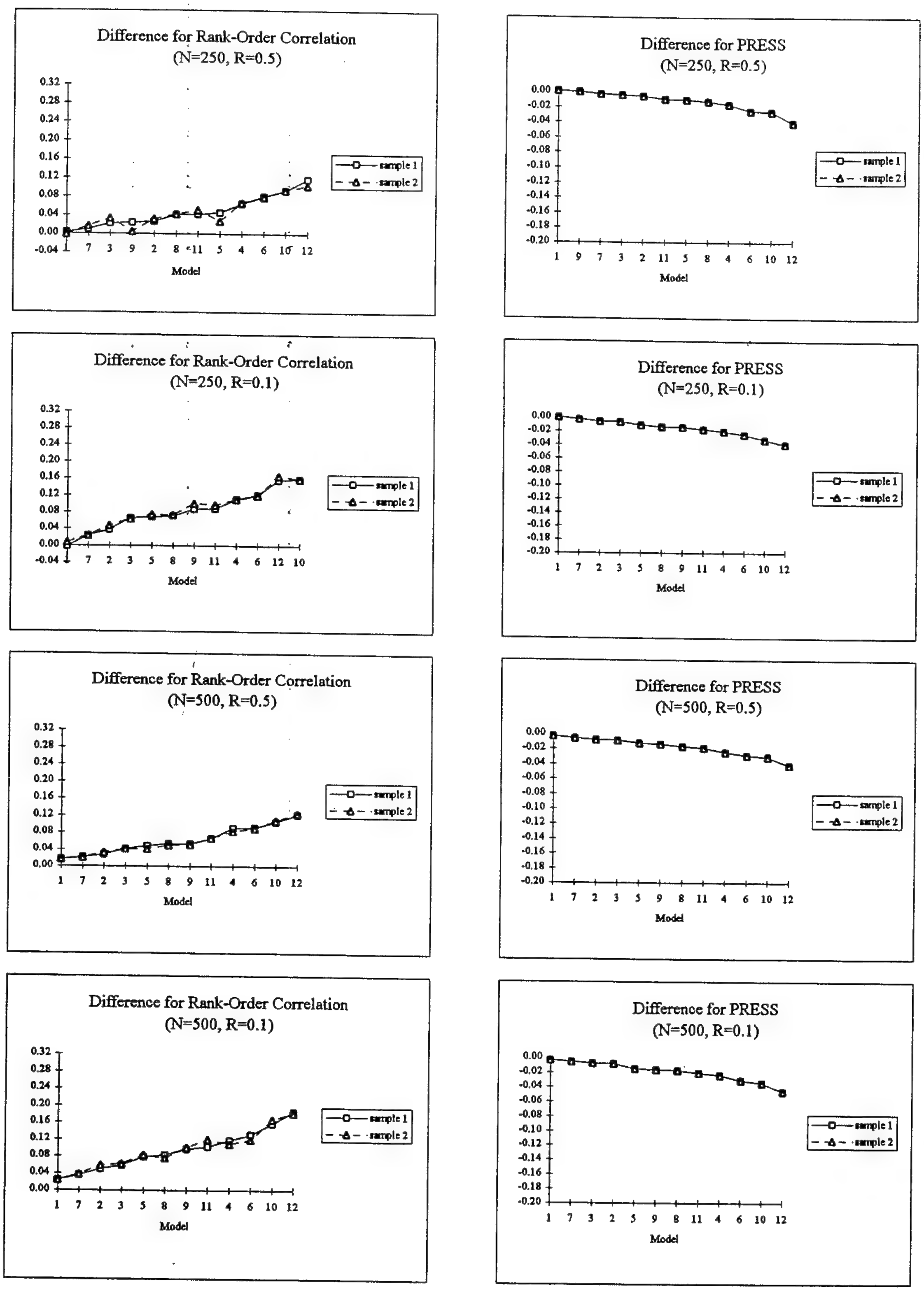


Figure 18
Plots of Difference for Cross-Validation of $g + 3\,s$

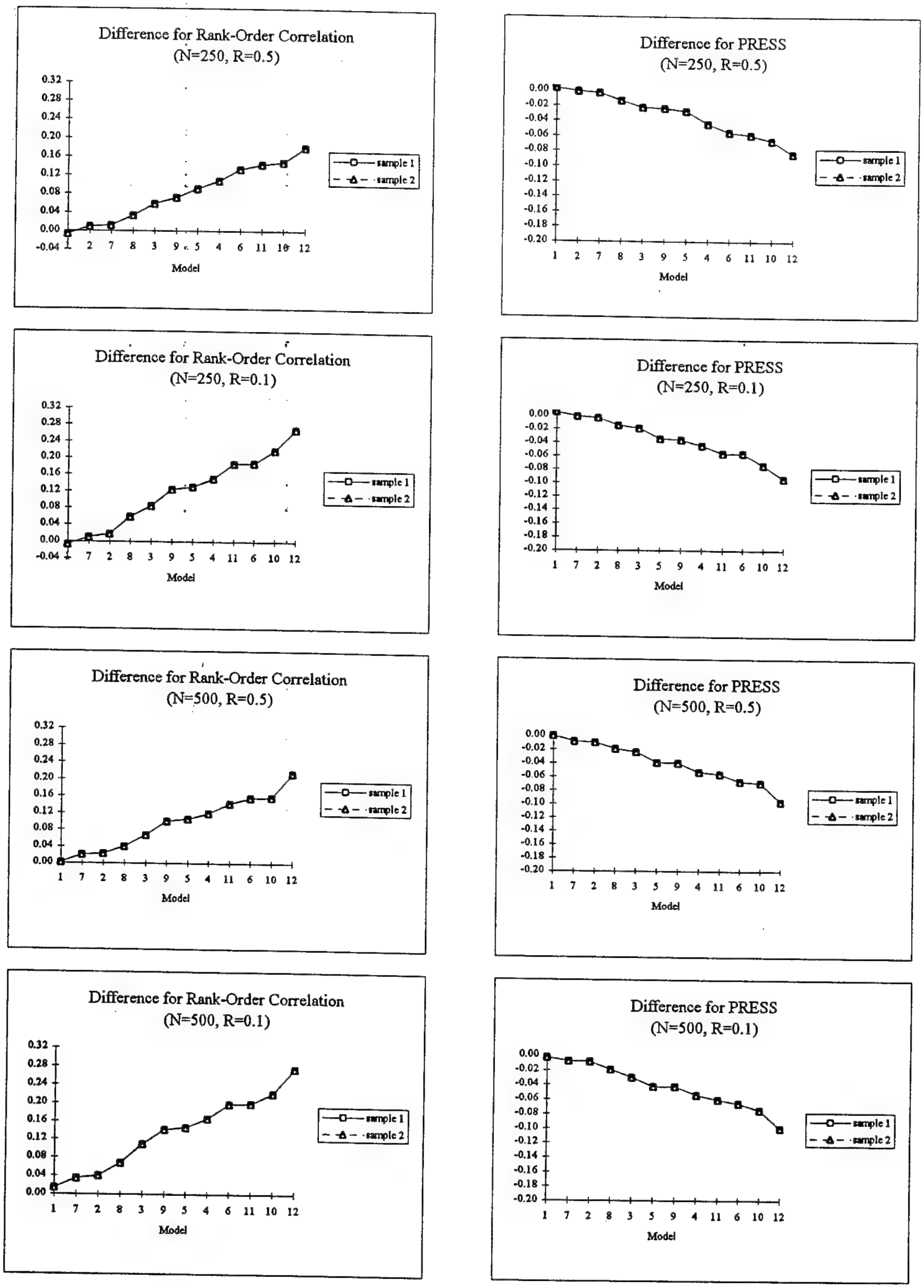
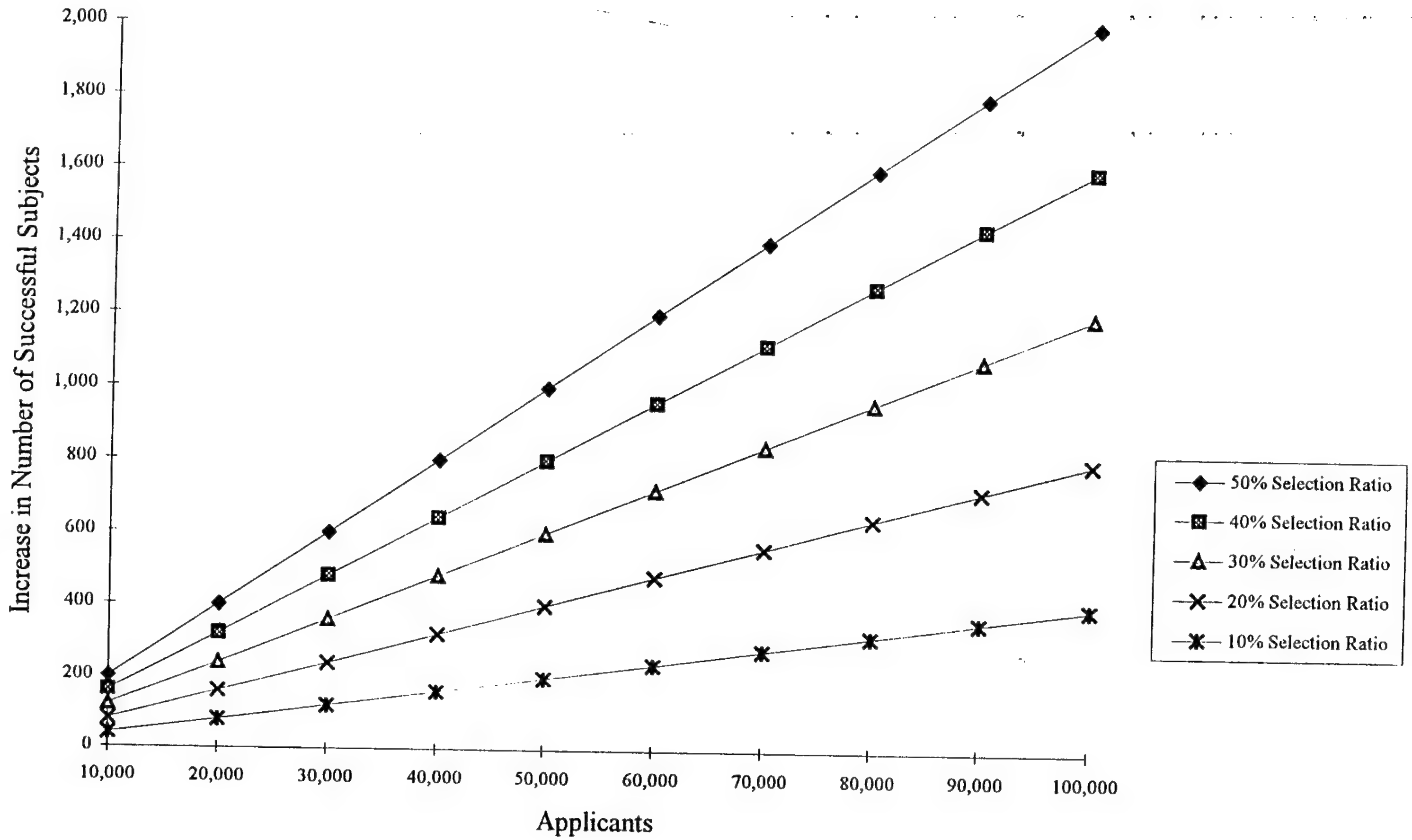


Figure 19
Plot for 4% Increase in the Rate of Sensitivity



Appendix A

The Factor Structure

Table A1

The Factor Structure for ASVAB and Project A Subtests

Subtest	Gf	Gc	Gv/Mech	Speed	Math	Psymotor
General Science	0.56	0.55	0.24	--	0.17	--
Arithmetic Reasoning	0.67	0.27	0.20	0.18	0.31	--
Word Knowledge	0.57	0.72	--	0.12	--	--
Paragraph Comprehension	0.56	0.54	--	0.22	--	--
Numerical Operations	0.42	--	--	0.72	0.16	--
Coding Speed	0.44	--	--	0.63	--	--
Auto & Shop Information	0.20	0.45	0.70	--	--	--
Mathematical Knowledge	0.68	0.13	--	0.16	0.57	0.08
Mechanical Comprehension	0.57	0.32	0.52	--	0.13	-0.05
Electronics Information	0.31	0.51	0.53	--	0.15	--
Assembling Objects	0.68	--	0.27	--	--	--
Figural Reasoning	0.77	--	0.17	--	--	--
Maze Test	0.56	--	0.29	--	--	-0.35
Object Rotation Test	0.48	--	0.31	--	--	-0.22
Orientation Test	0.63	--	0.28	--	--	--
Map Test	0.72	0.13	0.29	--	0.11	--
Target Tracking Test 1	-0.38	--	-0.35	--	--	0.30
Target Tracking Test 2	-0.40	--	-0.36	--	--	0.30
Target Identification, Time	-0.40	--	-0.27	--	--	0.41
Target Identification, Hits	0.23	--	--	--	--	--
Memory Search Test, Time	-0.17	0.10	-0.06	--	--	--
Memory Search Test, Hits	0.37	--	--	--	--	--

Note. From "The nature of the general factor in hierarchical models of the structure of cognitive abilities: alternative models tested on data from regular and experimental military enlistment tests," by J. E. Gustafsson and B. Muthen, 1994, UCLA Technical Report. p. 15. Adapted with permission of the author.

Table A2
The Factor Structure for ASVAB and ECAT Subtests

Subtest	Gf	Gc	Gv/Mech	Speed	Math	Psymotor
General Science	0.59	0.52	0.25	--	0.17	--
Arithmetic Reasoning	0.73	0.21	0.17	0.17	0.27	--
Word Knowledge	0.58	0.71	--	0.12	--	--
Paragraph Comprehension	0.54	0.55	--	0.24	--	--
Numerical Operations	0.41	--	--	0.71	0.16	--
Coding Speed	0.41	--	--	0.66	--	--
Auto & Shop Information	0.28	0.38	0.70	--	--	--
Mathematical Knowledge	0.68	0.12	0.49	0.18	0.57	--
Mechanical Comprehension	0.60	0.28	0.55	--	0.14	--
Electronics Information	0.38	0.44		--	0.14	--
Assembling Objects	0.72	--	0.29	--	--	--
Integrating Details	0.76	--	0.26	--	--	--
Sequential Memory	0.70	--	--	--	--	--
Figural Reasoning	0.76	--	0.11	--	--	--
Spatial Orientation	0.68	--	0.26	--	--	--
One-Hand Tracking	-0.44	--	-0.21	--	--	0.68
Two-Hand Tracking	-0.47	--	-0.27	--	--	0.72
Target Identification	-0.38	--	-0.16	--	--	0.23

Note. From "The nature of the general factor in hierarchical models of the structure of cognitive abilities: alternative models tested on data from regular and experimental military enlistment tests," by J. E. Gustafsson and B. Muthen, 1994, UCLA Technical Report, p. 12. Adapted with permission of the author.

Appendix B

Job Profile Description

Job 1 (Infantryman)

TABLE OF TOP BY NEW

TOP NEW

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	5	Total
0	62	71	85	86	87	391
	12.63	14.46	17.31	17.52	17.72	79.63
	15.86	18.16	21.74	21.99	22.25	
	62.00	71.00	85.00	86.00	95.60	
1	38	29	15	14	4	100
	7.74	5.91	3.05	2.85	0.81	20.37
	38.00	29.00	15.00	14.00	4.00	
	38.00	29.00	15.00	14.00	4.40	
Total	100	100	100	100	91	491
	20.37	20.37	20.37	20.37	18.53	100.00

TABLE OF TOP BY SIG

TOP SIG

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	5	Total
0	64	71	84	87	85	391
	13.03	14.46	17.11	17.72	17.31	79.63
	16.37	18.16	21.48	22.25	21.74	
	64.00	71.00	84.00	87.00	93.41	
1	36	29	16	13	6	100
	7.33	5.91	3.26	2.65	1.22	20.37
	36.00	29.00	16.00	13.00	6.00	
	36.00	29.00	16.00	13.00	6.59	
Total	100	100	100	100	91	491
	20.37	20.37	20.37	20.37	18.53	100.00

TABLE OF TOP BY GEN

TOP GEN

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	5	Total
0	62	70	87	88	84	391
	12.63	14.26	17.72	17.92	17.11	79.63
	15.86	17.90	22.25	22.51	21.48	
	62.00	70.00	87.00	88.00	92.31	
1	38	30	13	12	7	100
	7.74	6.11	2.65	2.44	1.43	20.37
	38.00	30.00	13.00	12.00	7.00	
	38.00	30.00	13.00	12.00	7.69	
Total	100	100	100	100	91	491
	20.37	20.37	20.37	20.37	18.53	100.00

Job 2 (Cannon Crewman)

TABLE OF TOP BY NEW

TOP NEW

Frequency Percent Row Pct Col Pct	1	2	3	4	5	Total
0	64	79	78	86	57	364
	13.79	17.03	16.81	18.53	12.28	78.45
	17.58	21.70	21.43	23.63	15.66	
	64.00	79.00	78.00	86.00	89.06	
1	36	21	22	14	7	100
	7.76	4.53	4.74	3.02	1.51	21.55
	36.00	21.00	22.00	14.00	7.00	
	36.00	21.00	22.00	14.00	10.94	
Total	100	100	100	100	64	464
	21.55	21.55	21.55	21.55	13.79	100.00

TABLE OF TOP BY SIG

TOP SIG

Frequency Percent Row Pct Col Pct	1	2	3	4	5	Total
0	65	78	75	88	58	364
	14.01	16.81	16.16	18.97	12.50	78.45
	17.86	21.43	20.60	24.18	15.93	
	65.00	78.00	75.00	88.00	90.63	
1	35	22	25	12	6	100
	7.54	4.74	5.39	2.59	1.29	21.55
	35.00	22.00	25.00	12.00	6.00	
	35.00	22.00	25.00	12.00	9.38	
Total	100	100	100	100	64	464
	21.55	21.55	21.55	21.55	13.79	100.00

TABLE OF TOP BY GEN

TOP GEN

Frequency Percent Row Pct Col Pct	1	2	3	4	5	Total
0	68	71	83	87	55	364
	14.66	15.30	17.89	18.75	11.85	78.45
	18.68	19.51	22.80	23.90	15.11	
	68.00	71.00	83.00	87.00	85.94	
1	32	29	17	13	9	100
	6.90	6.25	3.66	2.80	1.94	21.55
	32.00	29.00	17.00	13.00	9.00	
	32.00	29.00	17.00	13.00	14.06	
Total	100	100	100	100	64	464
	21.55	21.55	21.55	21.55	13.79	100.00

Job 3 (Tank Crewman)

TABLE OF TOP BY NEW

TOP / NEW Frequency Percent Row Pct Col Pct	1	2	3	4	Total
0	61 15.48 20.75 61.00	68 17.26 23.13 68.00	78 19.80 26.53 78.00	87 22.08 29.59 92.55	294 74.62
1	39 9.90 39.00 39.00	32 8.12 32.00 32.00	22 5.58 22.00 22.00	7 1.78 7.00 7.45	100 25.38
Total	100 25.38	100 25.38	100 25.38	94 23.86	394 100.00

TABLE OF TOP BY SIG

TOP / SIG Frequency Percent Row Pct Col Pct	1	2	3	4	Total
0	59 14.97 20.07 59.00	71 18.02 24.15 71.00	78 19.80 26.53 78.00	86 21.83 29.25 91.49	294 74.62
1	41 10.41 41.00 41.00	29 7.36 29.00 29.00	22 5.58 22.00 22.00	8 2.03 8.00 8.51	100 25.38
Total	100 25.38	100 25.38	100 25.38	94 23.86	394 100.00

TABLE OF TOP BY GEN

TOP / GEN Frequency Percent Row Pct Col Pct	1	2	3	4	Total
0	63 15.99 21.43 63.00	65 16.50 22.11 65.00	82 20.81 27.89 82.00	84 21.32 28.57 89.36	294 74.62
1	37 9.39 37.00 37.00	35 8.88 35.00 35.00	18 4.57 18.00 18.00	10 2.54 10.00 10.64	100 25.38
Total	100 25.38	100 25.38	100 25.38	94 23.86	394 100.00

Job 4 (Radio Operator)

TABLE OF TOP BY NEW

TOP / NEW

Frequency / Percent / Row Pct / Col Pct	1	2	3	Total
0	49	67	67	183
	17.50	23.93	23.93	65.36
	26.78	36.61	36.61	
	49.00	67.00	83.75	
1	51	33	13	97
	18.21	11.79	4.64	34.64
	52.58	34.02	13.40	
	51.00	33.00	16.25	
Total	100	100	80	280
	35.71	35.71	28.57	100.00

Frequency Missing = 9

TABLE OF TOP BY SIG

TOP / SIG

Frequency / Percent / Row Pct / Col Pct	1	2	3	Total
0	50	63	76	189
	17.30	21.80	26.30	65.40
	26.46	33.33	40.21	
	50.00	63.00	85.39	
1	50	37	13	100
	17.30	12.80	4.50	34.60
	50.00	37.00	13.00	
	50.00	37.00	14.61	
Total	100	100	89	289
	34.60	34.60	30.80	100.00

TABLE OF TOP BY GEN

TOP / GEN

Frequency / Percent / Row Pct / Col Pct	1	2	3	Total
0	45	70	74	189
	15.57	24.22	25.61	65.40
	23.81	37.04	39.15	
	45.00	70.00	83.15	
1	55	30	15	100
	19.03	10.38	5.19	34.60
	55.00	30.00	15.00	
	55.00	30.00	16.85	
Total	100	100	89	289
	34.60	34.60	30.80	100.00

Job 5 (Vehicle Mechanic)

TABLE OF TOP BY NEW

TOP: Frequency / Percent / Row Pct / Col Pct	NEW 1	2	3	4	5	Total
0	70 14.64 18.52 70.00	72 15.06 19.05 72.00	78 16.32 20.63 78.00	85 17.78 22.49 85.00	73 15.27 19.31 93.59	378 79.08
1	30 6.28 30.00 30.00	28 5.86 28.00 28.00	22 4.60 22.00 22.00	15 3.14 15.00 15.00	5 1.05 5.00 6.41	100 20.92
Total	100 20.92	100 20.92	100 20.92	100 20.92	78 16.32	478 100.00

TABLE OF TOP BY SIG

TOP: Frequency / Percent / Row Pct / Col Pct	SIG 1	2	3	4	5	Total
0	70 14.64 18.52 70.00	74 15.48 19.58 74.00	76 15.90 20.11 76.00	85 17.78 22.49 85.00	73 15.27 19.31 93.59	378 79.08
1	30 6.28 30.00 30.00	26 5.44 26.00 26.00	24 5.02 24.00 24.00	15 3.14 15.00 15.00	5 1.05 5.00 6.41	100 20.92
Total	100 20.92	100 20.92	100 20.92	100 20.92	78 16.32	478 100.00

TABLE OF TOP BY GEN

TOP: Frequency / Percent / Row Pct / Col Pct	GEN 1	2	3	4	5	Total
0	68 14.23 17.99 68.00	80 16.74 21.16 80.00	76 15.90 20.11 76.00	84 17.57 22.22 84.00	70 14.64 18.52 89.74	378 79.08
1	32 6.69 32.00 32.00	20 4.18 20.00 20.00	24 5.02 24.00 24.00	16 3.35 16.00 16.00	8 1.67 8.00 10.26	100 20.92
Total	100 20.92	100 20.92	100 20.92	100 20.92	78 16.32	478 100.00

Job 6 (Motor Transport)

TABLE OF TOP BY NEW

TOP / NEW (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	Total
0	57 11.24 14.00 57.00	77 15.19 18.92 77.00	84 16.57 20.64 84.00	88 17.36 21.62 88.00	101 19.92 24.82 94.39	407 80.28
1	43 8.48 43.00 43.00	23 4.54 23.00 23.00	16 3.16 16.00 16.00	12 2.37 12.00 12.00	6 1.18 6.00 5.61	100 19.72
Total	100 19.72	100 19.72	100 19.72	100 19.72	107 21.10	507 100.00

TABLE OF TOP BY SIG

TOP / SIG (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	Total
0	59 11.64 14.50 59.00	76 14.99 18.67 76.00	81 15.98 19.90 81.00	89 17.55 21.87 89.00	102 20.12 25.06 95.33	407 80.28
1	41 8.09 41.00 41.00	24 4.73 24.00 24.00	19 3.75 19.00 19.00	11 2.17 11.00 11.00	5 0.99 5.00 4.67	100 19.72
Total	100 19.72	100 19.72	100 19.72	100 19.72	107 21.10	507 100.00

TABLE OF TOP BY GEN

TOP / GEN (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	Total
0	65 12.82 15.97 65.00	70 13.81 17.20 70.00	88 17.36 21.62 88.00	82 16.17 20.15 82.00	102 20.12 25.06 95.33	407 80.28
1	35 6.90 35.00 35.00	30 5.92 30.00 30.00	12 2.37 12.00 12.00	18 3.55 18.00 18.00	5 0.99 5.00 4.67	100 19.72
Total	100 19.72	100 19.72	100 19.72	100 19.72	107 21.10	507 100.00

Job 7 (Administrative)

TABLE OF TOP BY NEW

TOP (Frequency / Percent / Row Pct / Col Pct)	NEW 1	2	3	4	Total
0	55	77	81	114	327
	12.88	18.03	18.97	26.70	76.58
	16.82	23.55	24.77	34.86	
	55.00	77.00	81.00	89.76	
1	45	23	19	13	100
	10.54	5.39	4.45	3.04	23.42
	45.00	23.00	19.00	13.00	
	45.00	23.00	19.00	10.24	
Total	100	100	100	127	427
	23.42	23.42	23.42	29.74	100.00

TABLE OF TOP BY SIG

TOP (Frequency / Percent / Row Pct / Col Pct)	SIG 1	2	3	4	Total
0	56	72	85	114	327
	13.11	16.86	19.91	26.70	76.58
	17.13	22.02	25.99	34.86	
	56.00	72.00	85.00	89.76	
1	44	28	15	13	100
	10.30	6.56	3.51	3.04	23.42
	44.00	28.00	15.00	13.00	
	44.00	28.00	15.00	10.24	
Total	100	100	100	127	427
	23.42	23.42	23.42	29.74	100.00

TABLE OF TOP BY GEN

TOP (Frequency / Percent / Row Pct / Col Pct)	GEN 1	2	3	4	Total
0	60	76	79	112	327
	14.05	17.80	18.50	26.23	76.58
	18.35	23.24	24.16	34.25	
	60.00	76.00	79.00	88.19	
1	40	24	21	15	100
	9.37	5.62	4.92	3.51	23.42
	40.00	24.00	21.00	15.00	
	40.00	24.00	21.00	11.81	
Total	100	100	100	127	427
	23.42	23.42	23.42	29.74	100.00

Job 8 (Medical)

TABLE OF TOP BY NEW

TOP / NEW (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	Total
0	59	70	80	83	292
	15.05	17.86	20.41	21.17	74.49
	20.21	23.97	27.40	28.42	
	59.00	70.00	80.00	90.22	
1	41	30	20	9	100
	10.46	7.65	5.10	2.30	25.51
	41.00	30.00	20.00	9.00	
	41.00	30.00	20.00	9.78	
Total	100	100	100	92	392
	25.51	25.51	25.51	23.47	100.00

TABLE OF TOP BY SIG

TOP / SIG (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	Total
0	60	69	81	82	292
	15.31	17.60	20.66	20.92	74.49
	20.55	23.63	27.74	28.08	
	60.00	69.00	81.00	89.13	
1	40	31	19	10	100
	10.20	7.91	4.85	2.55	25.51
	40.00	31.00	19.00	10.00	
	40.00	31.00	19.00	10.87	
Total	100	100	100	92	392
	25.51	25.51	25.51	23.47	100.00

TABLE OF TOP BY GEN

TOP / GEN (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	Total
0	59	71	79	83	292
	15.05	18.11	20.15	21.17	74.49
	20.21	24.32	27.05	28.42	
	59.00	71.00	79.00	90.22	
1	41	29	21	9	100
	10.46	7.40	5.36	2.30	25.51
	41.00	29.00	21.00	9.00	
	41.00	29.00	21.00	9.78	
Total	100	100	100	92	392
	25.51	25.51	25.51	23.47	100.00

Job 9 (Military Police)

TABLE OF TOP BY NEW

TOP NEW

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	5	6	Tot
0	64 10.72 12.88 64.00	77 12.90 15.49 77.00	87 14.57 17.51 87.00	90 15.08 18.11 90.00	86 14.41 17.30 86.00	93 15.58 18.71 95.88	497 83.25
1	36 6.03 36.00 36.00	23 3.85 23.00 23.00	13 2.18 13.00 13.00	10 1.68 10.00 10.00	14 2.35 14.00 14.00	4 0.67 4.00 4.12	100 16.75
Total	100 16.75	100 16.75	100 16.75	100 16.75	100 16.75	97 16.25	597 100.00

TABLE OF TOP BY SIG

TOP SIG

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	5	6	Tot
0	62 10.39 12.47 62.00	80 13.40 16.10 80.00	86 14.41 17.30 86.00	88 14.74 17.71 88.00	88 14.74 17.71 88.00	93 15.58 18.71 95.88	497 83.25
1	38 6.37 38.00 38.00	20 3.35 20.00 20.00	14 2.35 14.00 14.00	12 2.01 12.00 12.00	12 2.01 12.00 12.00	4 0.67 4.00 4.12	100 16.75
Total	100 16.75	100 16.75	100 16.75	100 16.75	100 16.75	97 16.25	597 100.00

TABLE OF TOP BY GEN

TOP GEN

Frequency / Percent / Row Pct / Col Pct	1	2	3	4	5	6	Tot
0	65 10.89 13.08 65.00	76 12.73 15.29 76.00	87 14.57 17.51 87.00	88 14.74 17.71 88.00	89 14.91 17.91 89.00	92 15.41 18.51 94.85	497 83.25
1	35 5.86 35.00 35.00	24 4.02 24.00 24.00	13 2.18 13.00 13.00	12 2.01 12.00 12.00	11 1.84 11.00 11.00	5 0.84 5.00 5.15	100 16.75
Total	100 16.75	100 16.75	100 16.75	100 16.75	100 16.75	97 16.25	597 100.00

Job 10 (Helicopter Mechanic)

TABLE OF TOP BY NEW

TOP / NEW (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	Total
0	61 13.90 17.99 61.00	71 16.17 20.94 71.00	81 18.45 23.89 81.00	87 19.82 25.66 87.00	39 8.88 11.50 100.00	339 77.22
1	39 8.88 39.00 39.00	29 6.61 29.00 29.00	19 4.33 19.00 19.00	13 2.96 13.00 13.00	0 0.00 0.00 0.00	100 22.78
Total	100 22.78	100 22.78	100 22.78	100 22.78	39 8.88	439 100.00

TABLE OF TOP BY SIG

TOP / SIG (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	Total
0	61 13.90 17.99 61.00	70 15.95 20.65 70.00	83 18.91 24.48 83.00	86 19.59 25.37 86.00	39 8.88 11.50 100.00	339 77.22
1	39 8.88 39.00 39.00	30 6.83 30.00 30.00	17 3.87 17.00 17.00	14 3.19 14.00 14.00	0 0.00 0.00 0.00	100 22.78
Total	100 22.78	100 22.78	100 22.78	100 22.78	39 8.88	439 100.00

TABLE OF TOP BY GEN

TOP / GEN (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	Total
0	68 15.49 20.06 68.00	72 16.40 21.24 72.00	75 17.08 22.12 75.00	88 20.05 25.96 88.00	36 8.20 10.62 92.31	339 77.22
1	32 7.29 32.00 32.00	28 6.38 28.00 28.00	25 5.69 25.00 25.00	12 2.73 12.00 12.00	3 0.68 3.00 7.69	100 22.78
Total	100 22.78	100 22.78	100 22.78	100 22.78	39 8.88	439 100.00

Job 11 (Automotive Mechanic)

TABLE OF TOP BY NEW

TOP / NEW (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	6	7	Total
0	70 10.09 11.78 70.00	77 11.10 12.96 77.00	78 11.24 13.13 78.00	91 13.11 15.32 91.00	89 12.82 14.98 89.00	96 13.83 16.16 96.00	93 13.40 15.66 98.94	594 85.59
1	30 4.32 30.00 30.00	23 3.31 23.00 23.00	22 3.17 22.00 22.00	9 1.30 9.00 9.00	11 1.59 11.00 11.00	4 0.58 4.00 4.00	1 0.14 1.00 1.06	100 14.41
Total	100 14.41	100 14.41	100 14.41	100 14.41	100 14.41	100 14.41	94 13.54	694 100.00

TABLE OF TOP BY SIG

TOP / SIG (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	6	7	Total
0	70 10.09 11.78 70.00	77 11.10 12.96 77.00	78 11.24 13.13 78.00	92 13.26 15.49 92.00	88 12.68 14.81 88.00	96 13.83 16.16 96.00	93 13.40 15.66 98.94	594 85.59
1	30 4.32 30.00 30.00	23 3.31 23.00 23.00	22 3.17 22.00 22.00	8 1.15 8.00 8.00	12 1.73 12.00 12.00	4 0.58 4.00 4.00	1 0.14 1.00 1.06	100 14.41
Total	100 14.41	100 14.41	100 14.41	100 14.41	100 14.41	100 14.41	94 13.54	694 100.00

TABLE OF TOP BY GEN

TOP / GEN (Frequency, Percent, Row Pct, Col Pct)	1	2	3	4	5	6	7	Total
0	72 10.37 12.12 72.00	77 11.10 12.96 77.00	86 12.39 14.48 86.00	91 13.11 15.32 91.00	86 12.39 14.48 86.00	92 13.26 15.49 92.00	90 12.97 15.15 95.74	594 85.59
1	28 4.03 28.00 28.00	23 3.31 23.00 23.00	14 2.02 14.00 14.00	9 1.30 9.00 9.00	14 2.02 14.00 14.00	8 1.15 8.00 8.00	4 0.58 4.00 4.26	100 14.41
Total	100 14.41	100 14.41	100 14.41	100 14.41	100 14.41	100 14.41	94 13.54	694 100.00

References

Acito, F., & Anderson, R. D. (1986). A simulation study of factor score indeterminacy. *Journal of Marketing Research, 23*, 111-118.

Alley, W. E. (1994). Recent advances in classification theory and practice. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 431-442). Hillsdale, NJ: Lawrence Erlbaum.

Allred, L. J. (1991). Alternatives to the validity coefficient for reporting the test-criterion relationship. In A. K. Wigdor and B. F. Green, Jr. (Eds.), *Performance assessment for the workplace: Vol. 2. Technical issues* (pp. 158-206). Washington, DC: National Academy Press.

Ash, P., & Kroeker, L. (1975). Personnel selection, classification, and placement. *Annual Review of Psychology, 26*, 481-507.

Bobko, P. (1983). An analysis of correlations corrected for attenuation and range restriction. *Journal of Applied Psychology, 68*(4), 584-589.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structure equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Borack, J. I. (1994). Estimating predictive validity when range restriction due to selection and attrition is present. *Military Psychology, 6*(3), 193-204.

Brodnick, R. J. & Ree, M. J. (1995). A structural model of academic performance, socioeconomic status, and Spearman's *g*. *Educational and Psychological Measurement, 55*(4), 583-594.

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65*(5), 543-558.

Callender, J. C., & Osburn, H. G. (1988). Unbiased estimation of sampling variance of correlations. *Journal of Applied Psychology, 73*(2), 312-315.

Campbell, C. H., Ford, P., Rumsey, M. G., Pulakos, E. D., Borman, W. C., Felker, D. B., De Vera, M. V., & Riegelhaupt, B. J. (1990). Development of multiple job performance measures in a representative sample of jobs. *Personnel Psychology, 43*, 277-300.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43*, 313-333.

Campbell, J. P., & Zook, L. M. (1992). *Building and retaining the career force: New procedures for accessing and assigning army enlisted personnel.* Technical Report 952, United States Army Research Institute for the Behavioral and Social Sciences.

Campbell, J. P., & Zook, L. M. (1991). *Improving the selection, classification, and utilization of army enlisted personnel: Final report on Project A*. Research Report 1597, United States Army Research Institute for the Behavioral and Social Sciences.

Carey, N. B., Hiatt, C. M., & Mayberry, P. W. (1994). *General intelligence and job performance*. Unpublished manuscript.

Carretta, T. R. (1990). Cross validation of experimental USAF pilot training performance models. *Military Psychology, 2*(4), 257-264.

Cornelius, E. T., Carron, T. J., & Collins, M. N. (1979). Job analysis models and job classification. *Personnel Psychology, 32*, 693-708.

Cornelius, E. T., Schmidt, F. L., & Carron, T. J. (1984). Job classification approaches and the implementation of validity generalization results. *Personnel Psychology, 37*, 247-260.

Cotter, K. L., & Raju, N. S. (1982). An evaluation of formula-based population squared cross-validity estimates and factor score estimates in prediction. *Educational and Psychological Measurement, 42*, 493-519.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions.* Urbana: University of Illinois Press.

Dobie, T., McFarland, K., & Long, N. (1986). Raw score and factor score multiple regression: An evaluative comparison. *Educational and Psychological Measurement, 46*, 337-347.

Dunnette, M. D. (1979). Personnel selection and classification systems. *Annual Review of Psychology, 30*, 477-525.

Dwyer, P. S. (1954). Solution of the personnel classification problem with the method of optimal regions. *Psychometrika, 18*(1), 11-26.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of the ASVAB for training grades. *Educational and Psychological Measurement, 52*, 721-725.

Gardner, H. (1985). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gibb, G. D. (1990). Initial validation of computer-based secondary selection system for student naval aviators. *Military Psychology, 2*(4), 205-219.

Gorsuch, R. L. (1980). Factor score reliabilities and domain validities. *Educational and Psychological Measurement, 40*, 895-897.

Green, B. F., Jr. (1976). On the factor score controversy. *Psychometrika, 41*(2), 263-266.

Green, B. F., Jr., & Mavor, A. S. (Eds.). (1994). *Modeling cost and performance for military enlistment*. Washington, DC: National Academy Press.

Green, B. F., Jr., & Wing, H. (Eds.). (1988). *Analysis of job performance measurement data: Report of a workshop*. Washington, DC: National Academy Press.

Guildford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 4.* (pp. 35-71). Hillsdale, NJ: Lawrence Erlbaum.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407-434.

Gustafsson, J.-E., & Muthén, B. O. (1994). *The nature of the general factor in hierarchical models of the structure of cognitive abilities: Alternative models tested on data from regular and experimental military enlistment tests*. Technical Report, UCLA.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: The University of Chicago Press.

Harris, C. W. (1967). On factors and factor scores. *Psychometrika, 32*(4), 363-379.

Harvet, R. J. (1986). Quantitative approaches to job classification: A review and critique. *Personnel Psychology, 39*, 267-289.

Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed.,. Vol. 2, 873-919). Palo Alto, CA: Consulting Psychologists Press.

Hogan, P. F. & Harris, D. A. (1994). Policy and management applications of the accession quality cost/performance trade-off model. In B. F. Green, Jr. & A. S. Mavor (Eds.),

Modeling cost and performance for military enlistment (pp. 129-157). Washington, DC: National Academy Press.

Horn, J. L. (1965). An empirical comparison of methods for estimating factor scores. *Educational and Psychological Measurement, 2*, 313-322.

Horn, J. L., & Miller, W. C. (1966). Evidence on problems in estimating common factor scores. *Educational and Psychological Measurement, 26*, 617-677.

Horst, P. (1954). The maximum expected correlation between two multiple-choice tests. *Psychometrika, 19*(2), 291-296.

Horst, P. (1956). Optimal test length for maximum differential prediction. *Psychometrika, 21*(1), 51-66.

Hsu, J. Y. (1991). *Sampling behavior in estimating predictive validity in the context of selection and latent variable modeling.* Unpublished doctoral dissertation, University of California, Los Angeles.

Hsu, J. Y. (1995). Sampling behaviour in estimating predictive validity in the context of selection and latent variable modeling: a Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 48*, 75-97.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology, 75*(3), 334-349.

Irving, P. G. (1993). On the use of personality measures in personnel selection. *Canadian Psychology, 34*(2), 208-214.

Jensen, A. R. (1986). *g*: Artifact or reality? *Journal of Vocational Behavior, 29* 301-331.

Jensen, A. R., & Weng, L. (1994). What is a good *g*? *Intelligence, 18*, 231-258.

Johnson, C. D., & Zeidner, J. (1991). *The economic benefits of predicting job performance: Vol. 2. Classification efficiency.* New York: Praeger.

Kukuk, C. R., & Baty, C. F. (1979). The misuse of multiple regression with composite scales obtained from factor scores. *Educational and Psychological Measurement, 39*, 277-290.

Kulik, C. T., Oldham, G. R., & Langner, P. H. (1988). Measurement of job characteristics: Comparison of the original and the revised job diagnostic survey. *Journal of Applied Psychology, 73*(3), 462-466.

Landy, F. J., & Shankster, L. J. (1994). Personnel selection and placement. *Annual Review of Psychology, 45*, 261-296.

Larson, G. E., & Wolfe, J. H. (1995). Validity results for *g* from an expanded test base. *Intelligence, 20*, 15-25.

Lastovicka, J. L., & Thamodaran, K. (1991). Common factor score estimates in multiple regression problems. *Journal of Marketing Research, 28*, 105-112.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* London: Butterworth.

Little, R. J., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Toronto: John Wiley & Sons, Inc.

Lohman, D. F. (1994). Implications of cognitive psychology for ability testing: Three critical assumption. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 145-172). Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M. (1980). *Application of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Maxwell, A. E. (1971). Estimating true scores and their reliabilities in the case of composite psychological tests. *British Journal of Mathematical and Statistical Psychology, 24*, 195-204.

McClelland, D. C. (1993). Intelligence in not the best predictor of job performance. *Current Directions in Psychological Science, 2*(1), 5-6.

McColy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*(4), 493-505.

McDonald, R. P., & Burr, E. J. (1967). A comparison of four methods of constructing factor scores. *Psychometrika, 32*(4), 381-401.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: the relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

Mendoza, J. L., & Reinhardt, R. N. (1991). Validity generalization procedures using sample-based estimates: A comparison of six procedures. *Psychological Bulletin, 110*(3), 596-610.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). (pp. 13-103). New York, NY: MacMillan.

Morris, J. D. (1979). A comparison of regression prediction accuracy on several types of factor scores. *American Educational Research Journal, 16*(1), 17-24.

Morris, J. D. (1980). The predictive accuracy of full-rank variables vs. various types of factor scores: Implications for test validation. *Educational and Psychological Measurement, 40*, 389-396.

Morris, J. D., & Guertin, W. H. (1977). The superiority of factor scores as predictors. *Journal of Experimental Education, 45*, 41-44.

Munson, J. M., & Posner, B. Z. (1980). Concurrent validation of two value inventories in predicting job classification and success for organizational personnel. *Journal of Applied Psychology, 65*(5), 536-542.

Muthén, B. O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*(4), 551-560.

Muthén, B. O. (1988). *LISCOMP: Analysis of linear structural equations with a comprehensive measurement model* (2nd ed.). Mooresville, IN: Scientific Software.

Muthén, B. O. (1989). Factor structure in groups selected on observed scores. *British Journal of Mathematical and Statistical Psychology, 42*, 81-90.

Muthén, B. O. (1994). *A note on predictive validity from a latent variable perspective.* Technical Report, UCLA.

Muthén, B. O., & Gustafsson, J.-E. (1995). *ASVAB-based job performance prediction and selection: Latent variable modeling versus regression analysis.* Technical Report, UCLA.

Muthén, B. O., & Hsu, J. Y. (1993). Selection and predictive validity with latent variable structures. *British Journal of Mathematical and Statistical Psychology, 46*, 255-271.

Muthén, B. O., & Jõreskog, K. G. (1983). Selectivity problems in quasi-experimental studies. *Evaluation Review, 7*(2), 139-174.

Muthén, B. O., Kao, C., & Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Application of a new IRT-based detection technique. *Journal of Educational Measurement, 28*, 1-22.

Muthén, B. O., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika, 42*, 431-462.

Muthén, B. O., Khoo, S. T., & Goff, G. N. (1994). *Multidimensional description of subgroup differences in mathematics achievement data from the 1992 National Assessment of Educational Progress*. Technical Report, UCLA.

Muthén, B. O., Hsu, J. Y., Carlstedt, B., & Mårdberg, B. (1994). *Predictive validity assessment of the Swedish military enlistment testing procedure using missing data and latent variable methods*. Technical Report, UCLA.

Nye, L. G. & Collins, W. E. (1993). Some personality and aptitude characteristics of air traffic control specialist trainees. *Aviation, space, and environmental Medicine, 64*, 711-716.

Olea, M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79*, 845-851.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd. ed.). Four Worth, Texas: Harcourt Brace Jovanovich, Inc.

Prediger, D. J. (1989). Ability differences across occupations: More than *g*. *Journal of Vocational Behavior, 34*(1), 1-27.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68*(3), 382-395.

Ree, M. J., & Carretta, T. R. (1995). Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement, 55*(2), 268-277.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior job knowledge in complex training performance. *Journal of Applied Psychology, 80*(6), 721-730.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. *Personnel Psychology*, *44*, 321-332.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, *1*, 86-89.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology*, *79*(4), 518-524.

Rumsey, M. G., Walker, C. B., & Harris, J. H. (Eds.). (1994). *Personnel selection and classification*. Hillsdale, NJ: Lawrence Erlbaum.

Sackett, P. R., Cornelius, E. T., & Carron, T. J. (1981). A comparison of global judgment vs. task oriented approaches to job classification. *Personnel Psychology*, *34*, 791-804.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 53-61.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*(5), 529-540.

Schmidt, F. L., Hunter, J. E., Croll, P. R., & McKenzie, R. C. (1983). Estimation of employment test validities by expert judgment. *Journal of Applied Psychology*, *68*(4), 590-601.

Schmidt, F. L., Law, K., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology*, *78*(1), 3-12.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology*, *43*, 627-670.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*(5), 607-620.

Scholarios, D. M., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology*, *79*(3), 412-424.

Schönemann, P. H., & Wang, M. (1972). Some new results on factor indeterminacy. *Psychometrika*, *37*(1), 61-91.

Short, L. M. (1990). *Factor score determinacy in a hierarchical multidimensional model.* Unpublished doctoral dissertation, University of California, Los Angeles.

Spearman, C. (1927). *The abilities of man.* London: MacMillan.

Stutzman, T. M. (1983). Within classification job differences. *Personnel Psychology, 36,* 503-516.

Tenopyr, M. L. (1982). Personnel selection and classification. *Annual Review of Psychology, 33,* 581-618.

Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703-742.

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29,* 332-339.

Tucker, L. R. (1971). Relations of factor score estimates to their use. *Psychometrika, 36*(4), 427-436.

Velicer, W. F. (1976). The relation between factor score estimates, image scores, and principal component scores. *Educational and Psychological Measurement, 36,* 149-159.

Vineberg. R., & Joyner, J. N. (1989). Evaluation of individual enlisted performance. In M. F. Wiskoff & G. M. Rampton (Eds.), *Military personnel measurement: Testing, assignment, evaluation* (pp. 169-200). New York, NY: Praeger Publishers.

Wackwitz, J. H., & Horn, J. L. (1971). On obtaining the best estimates of factor scores within an ideal simple structure. *Multivariate Behavioral Research, 6,* 389-408.

Weisberg, S. (1985). *Applied linear regression* (2nd. ed.). New York: John Wiley & Sons.

Williams, J. S. (1978). A definition for the common-factor analysis model and the elimination of problems of factor score indeterminacy. *Psychometrika, 43*(3), 293-306.

Wigdor, A. K., & Green, B. F. (Eds.). (1991a). *Performance assessment for the workplace: Vol. 1.* Washington, DC: National Academy Press.

Wigdor, A. K., & Green, B. F. (Eds.). (1991b). *Performance assessment for the workplace: Vol. 2 Technical issues.* Washington, DC: National Academy Press.

Wise, L. L., McHenry, J. J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology, 43*, 355-366.

Wolfe, J. H. (1994). *Can we increase the validity of military selection batteries?* Paper presented at the UCLA Conference on Methodological Problems in Prediction with Hierarchical Cognitive Models, Los Angeles.

Wolfe, J. H. , Alderton, D. L., Larson, G. E., & Held, J. D. (1995). *Incremental validity of enhanced computer administered testing (ECAT).* (TN-96-6). San Diego, CA: Navy Personnel Research and Development Center.

Zeidner, J. (1969). Improving human performance in an information processing system. *Occupational Psychology, 43*(1), 1-14.

Zeidner, J., & Johnson, C. D. (1991a). Classification efficiency and systems design. *Journal of the Washington Academy of Sciences, 81*(2), 110-128.

Zeidner, J., & Johnson, C. D. (1991b). *The economic benefits of predicting job performance: Vol. 1. Selection utility.* New York: Praeger.

Zeidner, J., & Johnson, C. D. (1991c). *The economic benefits of predicting job performance: Vol. 3. Estimating the gains of alternative policies.* New York: Praeger.